

Article

Auditing the Algorithmic Leviathan: A Tiered Accountability and Reporting Standards Framework for Democratic Public Administration

Haris Alibašić 

Department of Business Administration, Lewis Bear Jr. College of Business, University of West Florida,
Pensacola, FL 32514, USA; halibasic@uwf.edu

Abstract

The accelerating use of algorithmic systems in public administration exposes a standardization gap between technical AI assurance and institutional accountability. This article develops a tiered accountability and reporting standards framework for decisions made under public authority. A qualitative design science method combines comparative institutional analysis, structured absence analysis, and standards architecture design. The empirical basis comprises international standards, the United States Department of Government Efficiency (DOGE)–Treasury access episode as an institutional control precursor, Australia’s Robodebt scheme as an automated-decision failure, and public sector governance arrangements in Estonia, Singapore, Japan, South Korea, Canada, and the United States. A replicable coding protocol traces documented accountability gaps to five auditable primitives: provenance tracking, decision logging, role attribution, contestability, and post-deployment audit. The primitives are organized into minimum, heightened, and systemic/constitutional tiers according to material influence, rights and essential service effects, civil service integrity, institutional independence, and substitutive capacity. The article also specifies a Public Sector Algorithmic Accountability Statement (PAAS), cross-walks its ten disclosure fields to GRI 1, GRI 2, and GRI 3, and demonstrates its operation through a fully worked hypothetical benefits eligibility application. A tier assignment decision aid, an assurance cycle, and a cost–feasibility model support implementation, including in small institutions. The framework’s novelty lies not in claiming new lifecycle controls, but in consolidating those controls around the public decision configuration, escalating them according to public authority consequences, and joining internal evidence to comparable public reporting. The proposal shifts standardization from AI system certification alone toward auditable institutional answerability, governance sustainability, and constitutional integrity.



Academic Editor: Francesco Tajani

Received: 30 June 2026

Revised: 5 August 2026

Accepted: 11 August 2026

Published: 1 September 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: algorithmic accountability; artificial intelligence standards; public sector sustainability reporting; public administration; auditability; administrative justice; governance; hybrid intelligence; algorithmic auditing

1. Introduction

Artificial intelligence (AI) and algorithmic decision systems now influence procurement, benefit administration, workforce management, regulatory enforcement, tax administration, fraud screening, service triage, and policy analysis [1–3]. Their public value depends on more than predictive accuracy or operating efficiency. Decisions made under

public authority must remain lawful, attributable, reviewable, explainable to the affected person, and open to correction. The central standards question is therefore not only whether an AI system is technically well managed, but whether the institution exercising authority preserves an auditable chain from legal mandate to data, model output, human judgment, final decision, remedy, and public report.

The international standards ecosystem supplies substantial technical and organizational foundations. ISO/IEC 42001:2023 specifies requirements for organizational AI management systems; ISO/IEC 23894:2023 provides AI risk management guidance; the NIST AI Risk Management Framework organizes governance, mapping, measurement, and management; and IEEE 7000-2021 provides a process for addressing ethical concerns during system design [4–7]. The European Union Artificial Intelligence Act adds legally enforceable, risk-differentiated duties, including documentation, logging, human oversight, conformity assessment, and post-market monitoring for designated high-risk systems [8]. These instruments materially advance trustworthy AI. Their dominant regulatory object, however, remains the AI system, provider, deployer, organizational management system, or risk class. Public administration also requires decision-level and institution-level evidence: who possesses authority, which official accepts responsibility, what record reconstructs the decision, which forum hears a challenge, and how unresolved risks become visible outside the deploying organization.

This institutional layer is also a sustainability reporting problem. Public sector sustainability includes the durability of governing capacity, legal continuity, institutional memory, public trust, resource stewardship, and the ability of residents and oversight bodies to hold authorities answerable [9–13]. A public organization that reports environmental and financial indicators while omitting material algorithmic authority, access, appeal, vendor dependence, and audit findings presents an incomplete account of governance sustainability. Algorithmic accountability reporting should therefore connect operational control evidence to an established disclosure architecture rather than operate as an isolated technology inventory.

Prior scholarship already identifies many relevant accountability components. Wieringa maps the objects and forums of algorithmic accountability; Raji et al. specify an end-to-end internal audit process; and the Ada Lovelace Institute, AI Now Institute, and Open Government Partnership synthesize public sector accountability mechanisms [14–16]. Research on public accountability, administrative responsibility, and the principles-to-practice gap further explains why abstract commitments require inspectable institutional arrangements [17–20]. The present article does not relabel provenance, logging, role attribution, contestability, or auditing as newly discovered controls. Its narrower novelty is architectural: (i) it treats the public decision configuration—legal authority, system, data, officials, vendor, and review forum—as the unit of standardization; (ii) it introduces public authority tiers that escalate requirements when rights, essential services, civil service integrity, institutional independence, or constitutional relations are implicated; (iii) it operationalizes substitutive capacity as a continuity and anti-dependence control; and (iv) it joins the same evidence base to a standardized Public Sector Algorithmic Accountability Statement (PAAS).

Two contrasting failures motivate the design. The 2025–2026 DOGE–Treasury episode concerned rapid access to high-consequence federal payment systems. It is not presented as proof of an algorithmic decision failure. Its analytical role is narrower: it exposes the authorization, access control, role attribution, logging, and oversight preconditions that must exist before technologically empowered teams use or alter systems of public authority [21–23]. Australia’s Robodebt scheme supplies the missing automated-decision dimension. The scheme’s use of income averaging and automated debt-raising processes

produced a documented administrative failure in which legal authority, evidence quality, notice, explanation, review, and institutional responsibility did not operate as an integrated assurance chain [24,25]. Together, the cases distinguish an institutional control precursor from realized automated administrative harm.

The study addresses three research questions. RQ1 asks which auditable primitives are necessary to preserve institutional answerability when algorithmic systems materially influence public decisions. RQ2 asks how those primitives should escalate across levels of public authority consequence. RQ3 asks how internal control and assurance evidence should be converted into comparable public governance reporting. The article advances four *ex ante* design propositions. These propositions guide artifact construction and evaluation; they are normative claims instantiated in the framework, not statistical hypotheses and not empirically confirmed effects.

DP1. Decision-level institutional accountability should produce greater audit reconstructability than system-only assurance because it links authority, evidence, human responsibility, outcome, and review.

DP2. Risk-proportionate public authority tiering should reduce under-specification for rights-sensitive and systemic uses without imposing an identical assurance burden on low-consequence uses.

DP3. A PAAS aligned with recognized reporting principles should improve the completeness, comparability, and verifiability of public sector algorithmic governance disclosure.

DP4. Reuse of a common evidence base across management controls, impact assessment, audit, and reporting should improve implementation feasibility and limit duplicative documentation.

The objectives are to derive a replicable evidence-to-requirement trail, compare diverse operating mechanisms, specify a tier-assignment aid, demonstrate a completed PAAS, map PAAS to GRI reporting architecture, and examine administrative feasibility. The framework remains normative and requires field validation. It does not rank countries, estimate causal effects, or establish that formal rules are uniformly implemented. Section 2 presents the reproducible method; Sections 3–5 report the standards, negative-case, and comparative findings; Sections 6 and 7 develop the tiered framework and PAAS; Section 8 interprets the design propositions, implementation conditions, and limitations; and Section 9 concludes.

2. Materials and Methods

2.1. Research Design and Unit of Analysis

The study uses qualitative design science and comparative institutional analysis to build an artifact: a public sector accountability and reporting standards architecture [26–28]. Design science is appropriate because the research problem concerns how documented governance failures and existing institutional mechanisms should be translated into auditable requirements. The method combines four stages: problem diagnosis, structured case comparison, primitive derivation, and artifact evaluation through internal consistency, traceability, and a worked application.

The unit of analysis is the public decision configuration rather than the model alone. A configuration comprises the legal authority, institutional purpose, relevant data, algorithmic component, human roles, vendor relations, decision output, affected population, review forum, and recordkeeping environment. “Algorithmic system” is used broadly for computational systems that classify, rank, recommend, predict, generate, optimize, or execute an administrative action. “Material influence” exists when a system’s output affects eligibility, priority, timing, amount, intensity of scrutiny, allocation of a public resource, employment consequence, enforcement action, or the practical options available

to a decision-maker. A nominally advisory tool is therefore within scope when officials routinely follow its output or when workflow design makes deviation costly or exceptional.

2.2. Documentary Corpus and Selection Protocol

The documentary corpus was bounded to the material available through 30 June 2026. It includes: (a) international technical, management-system, audit, and reporting standards; (b) enacted legislation and official administrative guidance; (c) official audit, ombudsman, executive, legislative, and commission reports; and (d) peer-reviewed scholarship on public accountability, algorithmic auditing, administrative law, public sector digitalization, and design science. Official sources were preferred for claims about present legal or administrative arrangements. Commentary without a traceable primary basis was excluded from case coding.

The comparator-selection universe consisted of national public sector AI governance and operating frameworks among OECD members and widely documented digital government jurisdictions. Purposive, diverse-case selection applied four criteria: a documented public sector governance mechanism; primary source accessibility; relevance to at least one candidate accountability primitive; and variation in legal-administrative tradition or implementation instrument [26]. Estonia, Singapore, Japan, South Korea, Canada, and the United States were retained. The set is not geographically representative and does not imply superior national performance. It supplies variation across administrative law constraints, assurance testing, procurement guidance, framework legislation, mandatory impact assessment, and executive minimum practices. Canada and current U.S. Office of Management and Budget (OMB) guidance were added because they are directly analogous public sector operating regimes [29–32]. The review and coding protocol is summarized in Table 1.

Table 1. Replicable document review and coding protocol.

Stage	Decision Rule	Recorded Evidence	Reliability Safeguard
Corpus inclusion	Primary law, official policy/audit/reporting instrument, or peer-reviewed source available by 30 June 2026.	Citation, issuing body, date, jurisdiction, instrument type, scope.	Prefer primary sources; record supersession and effective dates.
Case selection	Documented governance mechanism or accountability failure relevant to public authority.	Selection criterion met and reason for inclusion.	Use diverse mechanisms; avoid ranking jurisdictions.
Coding	Score each accountability criterion 0–3.	0 absent; 1 principle; 2 procedural requirement; 3 auditable evidence requirement.	Apply one codebook to negative and reference cases.
Derivation	Retain a primitive only if necessary for answerability, inspectable, and supported across at least two evidence categories.	Gap → control → evidence artifact → primitive → tier/PAAS field.	Record contrary evidence and rival explanations.
Design evaluation	Test coherence, tier consistency, reporting crosswalk, and usability.	Decision aid, worked PAAS, cost model, unresolved ambiguities.	Treat findings as design adequacy, not causal confirmation.

2.3. Coding Criteria and Structured Absence Analysis

Each case or comparator instrument set was assessed against ten criteria: lawful authority; provenance and data quality; decision traceability; role attribution; notice and explanation; challenge and remedy; independent audit; public reporting; vendor dependence and continuity; and material resource use. Coding distinguished four levels: 0, absent; 1, expressed as a principle; 2, translated into a procedure or assigned duty; and 3, supported by an evidence artifact available for inspection. For the two negative cases, the scoring object was the contemporaneous accountability architecture at the decision or access point; later inquiries served as evidence of earlier presence or absence and did not raise the underlying score. For the reference jurisdictions, the scoring object was the formal mechanism in force by 30 June 2026, not national implementation performance. Where several instruments formed one case, the case-level score reflected the highest level demonstrably applicable to the public sector mechanism, subject to contrary or superseding evidence.

Level 3 required an inspectable artifact or an explicit duty to create one. Examples include an authorization record, versioned dataset lineage, decision log, signed human override, published impact assessment, appeal register, audit workpaper, corrective action record, contractual audit clause, or tested continuity plan. Supplementary Materials S1 provides the complete operational codebook (Table S1), case $\times$ criterion score matrix (Table S2), and criterion-level documentary anchors and rationales (Table S3). These materials expose every assigned score to replication and disagreement rather than treating the coding protocol as self-validating.

Structured absence analysis was applied to the DOGE–Treasury and Robodebt cases. The analysis asked what evidence would have been required to establish lawful authorization, bounded access, decision provenance, responsible officials, notice, contestability, and independent verification at the relevant moment. An absence was not coded merely because a source failed to mention an artifact. It was coded only when an authoritative inquiry identified a control weakness, when a required artifact was unavailable or insufficient for review, or when the institutional process left a material accountability question unresolved. This rule reduces the risk of treating documentary silence as proof of institutional failure.

Four safeguards address confirmation bias. First, the codebook was specified before translating findings into the five primitives. Second, the same criteria were applied to negative and positive cases. Third, contrary evidence and cases in which a formal mechanism remained incomplete were retained rather than suppressed. Fourth, a primitive was retained only when it was necessary to answer an accountability question, expressed as inspectable evidence, and triangulated across two or more categories of source. The study does not claim intercoder reliability because coding was conducted by one researcher; independent replication is part of the validation agenda.

2.4. Framework Construction and Evaluation

The derivation chain proceeded from observed governance problem to institutional control, required evidence artifact, accountability primitive, proportional tier, and PAAS disclosure. Candidate controls that depended only on broad ethical language were rejected. The tier architecture was evaluated against two constraints: no lower tier should omit the minimum evidence needed to reconstruct a materially influenced decision, and escalation should correspond to a distinct public authority consequence. PAAS was then crosswalked to GRI disclosure principles and governance/material topic disclosures [12,33,34]. A fully worked hypothetical application tests whether the framework identifies a tier, assigns roles, specifies evidence, discloses limitations, and exposes residual ambiguity without relying on confidential data.

Methodological limitations are explicit. The coding is qualitative; the tiers are normative rather than calibrated risk scores; official documents do not establish uniform implementation; and the worked example demonstrates logical feasibility rather than measured organizational performance. The framework is intended for pilot testing, inter-rater evaluation, and jurisdiction-specific legal adaptation.

3. Results: The Institutional Standardization and Reporting Gap

3.1. Existing Standards and Operating Frameworks

The review identified five partially overlapping governance layers. Technical and management system standards organize risk and controls; risk-based law assigns obligations to system classes and actors; public sector operating guidance specifies impact assessments and minimum practices; administrative law governs lawful decisions and remedies; and sustainability/governance reporting establishes disclosure principles. No reviewed instrument fully integrates all five layers around a reconstructable public decision.

ISO/IEC 42001, ISO/IEC 23894, NIST AI RMF, IEEE 7000, and ISO 19011 support governance, risk treatment, values integration, and audit practice [4–7,35]. The EU AI Act adds enforceable lifecycle duties for designated uses [8]. These instruments support much of the technical evidence required by the proposed framework. They do not, on their own, identify the public official responsible for a particular administrative outcome, establish a legally sufficient appeal route, or require a consolidated governance disclosure tied to material public decisions.

Canada's Directive on Automated Decision-Making and mandatory Algorithmic Impact Assessment (AIA) move closer to an operating public sector regime. The AIA organizes 65 risk questions and 41 mitigation questions across project, system, algorithm, decision, impact, and data factors, produces one of four impact levels, and supports publication of completed assessments [29,30]. In the United States, OMB Memorandum M-25-21 replaced M-24-10 and requires governance structures, inventories, and minimum practices for high-impact AI, including discontinuation where risks cannot be adequately mitigated. OMB Memorandum M-25-22 replaced M-24-18 and addresses acquisition, competition, portability, interoperability, vendor sourcing, performance, and cross-functional oversight [31,32]. These regimes demonstrate that impact classification and minimum practices are administratively feasible. Their focus remains principally internal governance, assessment, or acquisition; PAAS adds a common decision-level public reporting module.

GRI 1 establishes reporting principles of accuracy, balance, clarity, comparability, completeness, sustainability context, timeliness, and verifiability [12]. GRI 2 addresses assurance, governance responsibility, policy commitments, remediation, advice and concern mechanisms, and legal compliance; GRI 3 addresses material topic determination and management [33,34]. These standards offer a mature disclosure architecture, but they do not specify algorithmic decision content. PAAS fills that content gap while preserving GRI's reporting logic. Table 2 summarizes the five governance layers, their contributions, and the residual institutional gaps.

Table 2. Standards layers, strengths, and residual institutional gap.

Layer	Representative Instruments	Primary Contribution	Residual Gap Addressed by This Study
AI management and risk	ISO/IEC 42001; ISO/IEC 23894; NIST AI RMF	Lifecycle governance, risk identification, management controls.	Decision-specific authority, attributable official, remedy, and public disclosure.
Ethics and audit process	IEEE 7000; ISO 19011	Values integration and management system audit discipline.	Public authority tiering and mandatory evidence for affected persons and oversight forums.
Risk-based law	EU AI Act; South Korea framework legislation	Actor duties and differentiated legal obligations.	Institution-centered escalation for systemic or constitutional consequences.
Public sector operating regimes	Canada AIA; OMB M-25-21 and M-25-22	Impact assessment, high-impact minimum practices, inventories, acquisition controls.	Common external disclosure module tied to decision records and remedies.
Governance reporting	GRI 1, GRI 2, GRI 3	Quality principles, governance responsibility, remediation, material topic reporting.	Algorithmic accountability content and assurance evidence.

3.2. Novelty Boundary and Design Gap

The literature review confirms a substantial overlap between the five primitives and existing taxonomies [14–16,18,19]. That overlap is desirable because an auditable standard should consolidate mature controls rather than depend on novel terminology. The unresolved problem is the architecture joining those controls. System-centered frameworks may document a model while leaving the administrative decision and review forum fragmented. Internal audit frameworks may identify lifecycle checkpoints without establishing a public reporting duty. Public sector accountability inventories may describe mechanisms without a proportional tier for institutional independence or a shared evidence model for sustainability reporting.

The design gap therefore has three dimensions. First, the object of accountability must include the institution and decision configuration, not only the system. Second, proportionality must respond to the kind of public authority exercised, not only model characteristics or sector. Third, disclosure must carry evidence from authorization, operation, challenge, audit, and remediation into a comparable report. These dimensions define the framework's contribution and prevent claims that established controls are original merely because they are assembled under new labels.

4. Results from the Negative Cases

4.1. DOGE–Treasury as an Institutional Control Precursor

Executive Order 14158 established the United States DOGE Service and agency-level teams to modernize federal technology and improve governmental efficiency [21]. Litigation and congressional analysis subsequently raised questions about access to protected agency records [22]. On 28 April 2026, the U.S. Government Accountability Office reported that a DOGE team employee had received access to three Bureau of the Fiscal Service payment systems and that the bureau had not fully implemented three of four selected control areas relating to access, confidentiality, integrity, and usage monitoring [23].

The episode does not establish that a deployed algorithm made an unlawful payment decision. It is relevant because algorithmic accountability depends on prior institutional controls. A system's output is not auditable when access authority is unclear, users are not fully bounded to roles, logs do not support reconstruction, or independent review lacks timely evidence. Rapid technical access to a high-consequence administrative platform therefore functions as a precursor case: it reveals the controls required before data extraction, model training, optimization, prioritization, code change, or automated execution occur.

Four design implications follow. Authorization must be recorded at the level of user, purpose, system, dataset, duration, and permitted action. Access and change logs must be immutable enough for independent reconstruction. Technical personnel and public officials must have differentiated, named responsibilities. Oversight bodies must receive evidence before high-consequence access becomes irreversible, rather than reconstruct authority only after controversy.

4.2. Robodebt as an Automated Administrative Failure

Australia's Robodebt scheme provides a documented automated-decision failure with direct administrative consequences. The scheme used income averaging and automated processes in identifying and raising social security debts. The Commonwealth Ombudsman's 2017 investigation identified problems involving fairness, transparency, usability, communications, and complaint handling [25]. The Royal Commission's 2023 report examined the scheme's lawfulness, design, implementation, effects, and institutional responsibility [24].

Robodebt shows why technical operation and institutional legality cannot be separated. The Royal Commission found that the scheme proceeded without a lawful basis, that legal warnings were not acted upon, and that departmental conduct at times misled or impeded the Ombudsman's scrutiny [24]. The central failure was not merely that an automated process produced erroneous debts; it was that an unlawful design continued through organizational arrangements that obscured, displaced, or muted responsibility and contrary advice. Data transformations that appear administratively efficient may be unsuitable evidence for a legally consequential individual decision. Notices may disclose a result without giving an affected person enough information to understand the evidentiary basis. Human review may exist formally while workload, workflow defaults, or burden shifting weaken meaningful contestability.

The resulting design requirements are direct: provenance must describe not only source data but transformations and inferential limits; the decision log must connect each affected case to versions and inputs; role attribution must name authority for policy, system, legal interpretation, and final action; contestability must provide timely notice, accessible evidence, human reconsideration, and suspension or correction where appropriate; and post-deployment audit must examine legality, distributional effects, error, complaints, overrides, and remediation.

The architecture's contribution is visibility and earlier contestability, not guaranteed prevention. PAAS, the five primitives, and tiered assurance would require contemporaneous legal authority, contrary advice, data transformations, named approvals, decision records, challenge outcomes, and unresolved audit findings to be preserved and exposed to an independent forum. Those requirements would make contradictions between legal advice, system design, operation, and public justification more difficult to suppress and easier to challenge before harm compounds. They would not prevent officials determined to pursue an unlawful design, disregard evidence, conceal advice, or disable oversight; prevention still depends on officials and institutions acting on the evidence.

4.3. Derivation Trail

The derivation trail answers RQ1. The five primitives are not inferred from a single incident. Each is supported by the negative cases, accountability theory, audit literature, and at least one comparator mechanism. Their value lies in the evidence they require: an oversight forum should be able to inspect who had authority, what information entered the decision, what occurred, who was responsible, how the affected person obtained review, and what independent assurance found after deployment. Table 3 summarizes the derivation from each documented problem or mechanism to its required control, evidence artifact, and accountability primitive.

Table 3. Derivation from documented problem to auditable primitive.

Observed Problem or Mechanism	Required Control	Inspectable Evidence Artifact	Primitive/Downstream Use
Rapid access to high-consequence systems; incomplete access and usage controls.	Purpose-bound authorization and data lineage.	Signed authorization; access register; source and transformation record.	Provenance tracking; Tier III access controls; PAAS fields 2–4.
Weak reconstruction of who viewed, changed, recommended, or executed.	Event and decision logging with version identifiers.	Immutable access/change log; case-level decision record.	Decision logging; all tiers; PAAS field 5.
Known legal concerns were not acted upon, and oversight did not receive the full legal position.	Documented legal release gate, protected dissent and escalation, and direct oversight access.	Signed legal position; advice register; dissent or escalation record; disclosure log.	Role attribution and post-deployment audit; Tier II/III; PAAS fields 3, 6, and 9.
Insufficient notice, explanation, and practical ability to dispute debt.	Accessible notice, evidence access, human reconsideration, and remedy.	Notice template; appeal register; response time; correction record.	Contestability; Tier II/III; PAAS field 8.
Control problems become visible after operation; weak learning from complaints.	Independent, risk-based post-deployment review and corrective action.	Audit report; incident trend; validation record; closure evidence.	Post-deployment audit; PAAS fields 7 and 9.
Vendor or team dependence may prevent lawful fallback.	Tested substitutive capacity and exit arrangements.	Fallback procedure; portability test; recovery exercise; exit clause.	Tier III continuity; PAAS field 10.

Note: These controls strengthen visibility, traceability, and contestability; they do not guarantee prevention when officials deliberately pursue illegality, suppress contrary advice, or disable oversight.

5. Comparative Institutional Reference Cases

5.1. Selection Logic and Case Contributions

The six reference jurisdictions were selected for mechanism diversity rather than national ranking. Estonia links digital administration to administrative law constraints; Singapore supplies testing and assurance tools; Japan emphasizes human-centric principles, public procurement, and adaptive guidance; South Korea illustrates framework legislation and high-impact differentiation; Canada operates a mandatory AIA with four impact levels; and current U.S. OMB guidance combines high-impact minimum practices with acquisition, interoperability, and vendor controls [26–32,36–48]. This selection extends beyond an Estonia-plus-Asia comparison and includes the two most directly analogous North American public sector operating regimes. Table 4 summarizes the selection basis, mechanism, design lesson, and limitation for each comparator.

Table 4. Comparator selection and analytical contribution.

Jurisdiction	Selection Basis	Mechanism Examined	Design Lesson and Limitation
Estonia	Mature digital administration within codified administrative law.	Administrative Procedure Act, Taxation Act, scholarship on algorithmic administrative acts.	Digital process remains subject to competence, reasons, procedure, and review; formal law does not prove implementation quality [41–43].
Singapore	Operational assurance and governance tooling in a high-capacity digital state.	Model AI Governance Framework and AI Verify.	Testing artifacts and governance roles support evidence-based assurance; transfer requires adaptation to public law remedies [39,44–46].
Japan	Human-centric principles plus evolving public procurement/use guidance.	Social Principles, AI legislation, generative AI procurement and utilization guideline.	Versioned guidance links innovation, procurement, responsibility, and review; soft-law reliance may produce variable uptake [40,47,48].
South Korea	National framework legislation with differentiated high-impact obligations.	AI Framework Act and institutional governance arrangements.	Risk differentiation and state capacity support tiering; implementation evidence is still developing [36,37].
Canada	Mandatory public sector algorithmic impact assessment.	Directive on Automated Decision-Making and published AIA.	Four impact levels and mitigation questions show operational tiering; assessment quality still depends on evidence, oversight, and updates [29,30].
United States	Current federal high-impact governance and acquisition policy.	OMB M-25-21 and M-25-22.	Inventories, minimum practices, discontinuation, portability, interoperability, and vendor controls support implementation; executive guidance remains context-dependent [31,32].

5.2. Cross-Case Mechanisms and Evidentiary Implications

The six mechanisms contribute distinct elements to a shared evidence architecture. Estonia anchors digital administration in competence, reasons, notification, and review [41–43]. Singapore links internal governance to test artifacts, thresholds, responsible roles, and corrective action, illustrating the difference between a principle and assurance-ready evidence [39,44–46]. Japan joins human-centric principles with versioned procurement and utilization guidance, emphasizing documentation access, material change notice, incident cooperation, portability, security, and exit support [40,47,48]. South Korea contributes statutory differentiation for high-impact AI and shows that proportional duties require institutional review capacity [36,37]. Canada operationalizes four impact levels through a mandatory AIA whose questions, mitigation duties, and publication requirements support repeatable classification [29,30]. U.S. OMB guidance adds high-impact minimum practices, inventory and governance duties, discontinuation where risks remain inadequately mitigated, and acquisition controls for portability, interoperability, performance, and vendor dependence [31,32].

Across the cases, the transferable unit is the mechanism and its evidence artifact, not a claim of national superiority or uniform implementation. Formal legality without operational records, voluntary testing without public law remedies, legislation without review capacity, and self-assessment without independent challenge each remain incom-

plete. The comparative contribution lies in combining legal authority, test evidence, impact classification, procurement discipline, continuity, contestability, and public disclosure while retaining those limitations.

5.3. Comparative Synthesis and Transferability

Taken together, the cases support a complementary rather than competitive architecture: legality sets the boundary; testing converts claims into evidence; impact assessment and differentiated duties calibrate assurance; procurement and portability protect continuity; and reporting exposes material limitations. No case alone supplies the full chain. Their combined value lies in showing how authorization, operation, challenge, audit, remediation, and disclosure may reuse the same evidence while preserving public law accountability.

The comparative findings also clarify transferability. Administrative law form, constitutional review, executive authority, audit institutions, and procurement rules differ. The framework therefore specifies functions rather than prescribing one national institution. A review forum may be a court, tribunal, ombudsman, auditor-general, inspector general, parliamentary committee, data protection authority, regulator, or other body with sufficient independence and access. What matters is that the forum receives timely evidence, is able to question responsible actors, and has the power to require correction or cessation.

6. Tiered Accountability and Reporting Standards Architecture

6.1. Five Accountability Primitives

Provenance tracking records the lawful origin, authority, transformation, permitted use, and retention of data, rules, models, prompts, external knowledge sources, and material system components. It should distinguish observed facts from inferred variables and identify limitations relevant to administrative use.

Decision logging creates a reconstructable record of system version, material configuration, relevant inputs, output, confidence or uncertainty where meaningful, human action, override, timestamp, and final disposition. Logging should be proportionate and privacy-preserving, but it must preserve enough evidence to review a contested decision and detect aggregate patterns.

Role attribution assigns named institutional responsibility for legal authority, policy ownership, data stewardship, model approval, operational use, security, human review, incident response, audit, and remediation. Vendor responsibility supplements rather than replaces public responsibility.

Contestability provides advance or contemporaneous notice where appropriate, understandable reasons, access to material evidence, an accessible channel for questions or challenge, meaningful human reconsideration, time limits, protection against retaliation, and documented remedy. A nominal appeal that reproduces the same automated logic without independent judgment does not satisfy the primitive.

The post-deployment audit evaluates continuing legality, data and model performance, distributional effects, security, complaints, overrides, incidents, vendor changes, and remediation. Audit intensity should match the tier, but every in-scope system requires periodic review and a change-triggered reassessment.

6.2. Tier Assignment Decision Aid

Tier assignment follows the highest applicable consequence rather than an average score. Step 1 asks whether the system materially influences a public decision. If not, the institution records the rationale for exclusion and monitors scope drift. If yes, Tier I applies at minimum. Step 2 asks whether the system affects an individual's rights, legal status,

benefit, essential service, enforcement exposure, employment, civil service protection, or similarly consequential opportunity. If yes, Tier II applies. Step 3 asks whether the system affects institutional independence, democratic or judicial processes, oversight access, nationwide or critical fiscal administration, branch relations, emergency powers, mass personnel action, or a dependency without tested substitution. If yes, Tier III applies. Aggregated low-stakes outputs may escalate when their combined effect becomes material. Table 5 summarizes the cumulative assignment rules and evidence requirements.

Table 5. Integrated tier assignment decision aid and cumulative requirements.

Trigger or Edge Case	Assignment Rule	Required Evidence and Cumulative Requirements
No material influence on a public decision	Outside PAAS scope; retain a documented exclusion and monitor use.	Purpose, users, prohibited uses, review date, and evidence that output does not affect decision options.
Any material influence without a higher-tier trigger	Tier I at minimum.	Inventory; legal and policy basis; provenance; version and decision logs; named owner and reviewer; correction and complaint route; periodic and change-triggered review; incident escalation; material PAAS fields.
Rights, legal status, benefit, essential service, tax, enforcement, employment, or civil service action	Tier II at minimum.	Tier I plus legal review and impact assessment; representative validation; notice, reasons, evidence access, human reconsideration, time-bound appeal and remedy; independent validation and recurring audit; full PAAS and material incident disclosure.
Institutional independence, democratic or judicial process, oversight access, critical fiscal administration, emergency power, mass workforce action, or an untested critical dependency	Tier III.	Tier II plus highest-body authorization; separation of duties; independent institutional review; external oversight access; protected records; cessation thresholds; pre/post-deployment PAAS; tested substitutive capacity and recovery.
Advisory score routinely followed	Treat as material influence; apply the highest applicable tier.	Override rate, workflow defaults, user interviews, and decision correlation; then apply the cumulative evidence for Tier I, II, or III.
Fraud flag triggers investigation but not final sanction	Tier II because scrutiny and burden are consequential.	Threshold rationale, false-positive monitoring, investigation safeguards, notice, human review, appeal, and remedy.
Generative drafting assistant using case files	Classify by downstream influence; exclude only with documented non-influence.	Use restrictions, verification duty, confidentiality controls, sampled review, and evidence that output does not structure the final decision.
Critical vendor lacks tested replacement	Escalate to Tier III until substitution is demonstrated.	Portability, data/model export, fallback staffing, recovery exercise, audit and exit clauses, and continuity status.
Material change, new data, scope expansion, or complaint trend	Reassess; the highest resulting tier governs.	Change log, updated impact assessment, approval, validation, public update, and all cumulative controls for the revised tier.

6.3. Tier I—Minimum Institutional Accountability

Tier I applies to every system that materially influences a public decision but does not meet a higher-tier trigger. Required evidence includes a system and purpose record; legal and policy basis; data and model provenance; version and decision logs; named owner and

reviewer; basic notice where a person is affected; correction and complaint route; periodic performance and security review; incident escalation; and inclusion in the institution's algorithmic inventory. Tier I is not a low-quality tier. It is the universal accountability floor.

6.4. Tier II—Heightened Rights and Administrative Justice Accountability

Tier II applies when a system affects rights, legal status, essential services, benefits, licensing, enforcement, tax, employment, civil service protections, or comparable high-consequence opportunities. It adds a documented impact assessment; legal review; representative validation; subgroup and error analysis where lawful and relevant; accessible reasons; meaningful human reconsideration; appeal time limits; monitoring of overrides, complaints, and remedies; independent pre-deployment validation; periodic external or functionally independent audit; and prompt public disclosure of material incidents and corrective action. Automated final action without adequate human review requires specific legal authority and stronger evidentiary justification.

6.5. Tier III—Systemic/Constitutional Accountability

Tier III applies when algorithmic capacity may alter institutional independence, oversight access, democratic or judicial processes, branch relations, nationwide or critical fiscal administration, emergency powers, mass public workforce action, or continuity of an essential state function. It adds authorization by the highest accountable governance body; documented separation of duties; independent legal and constitutional review; direct access for external oversight; staged deployment; enhanced security and access control; real-time or near-real-time anomaly monitoring where justified; public reporting before and after material deployment; cessation thresholds; protected records; and tested substitutive capacity.

“Constitutional” is functional rather than limited to jurisdictions with a codified constitution or strong-form judicial review. In parliamentary and other systems, the tier applies to functions that preserve lawful allocation of authority, ministerial and civil service responsibility, legislative scrutiny, audit independence, electoral integrity, judicial or tribunal review, and protected oversight. Equivalent forums may include parliament, an auditor-general, ombudsman, independent regulator, tribunal, court, or data-protection authority.

Substitutive capacity is the institution's demonstrated ability to maintain lawful service and accountability when a system, vendor, model, or specialized team becomes unavailable or unsuitable. It is evidenced through a current fallback procedure; trained staff; accessible documentation; data and model output portability; defined maximum tolerable outage and recovery objective; a tested manual or alternative workflow; vendor exit provisions; and a recorded exercise showing that the fallback preserves legal review and service continuity. A paper plan without a rehearsal does not demonstrate substitutive capacity. Figure 1 depicts the relationship among public authority, the five accountability primitives, the proportional tiers, PAAS reporting, and assurance feedback.

Institutional Accountability and Reporting Architecture

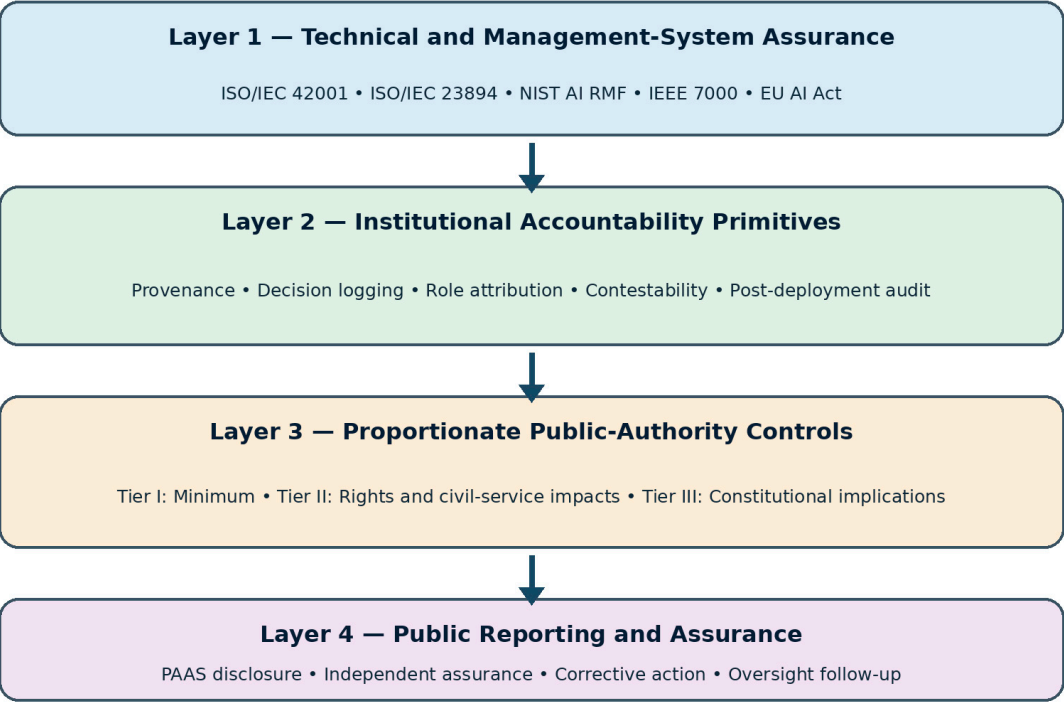


Figure 1. Relationship among public authority, accountability primitives, proportional tiers, PAAS reporting, and assurance feedback.

7. Public Sector Algorithmic Accountability Statement

7.1. Purpose and Ten Disclosure Fields

PAAS is a standardized governance disclosure for systems that materially influence public decisions. It is designed for an annual governance, sustainability, integrated, or equivalent public report, with interim updates after material change or incident. PAAS is not a model card, impact assessment, procurement file, or audit report. It summarizes the material institutional facts from those records and points to the underlying evidence.

The ten fields are: (1) system identity and public purpose; (2) decision scope, materiality, affected population, and assigned tier; (3) legal authority and policy basis; (4) data provenance, quality, and permitted use; (5) model, configuration, change, and decision logging; (6) named institutional roles and human review; (7) performance, impacts, limitations, and resource use; (8) notice, contestability, appeals, and remedies; (9) audit, assurance, incidents, findings, and corrective action; and (10) vendor dependencies, portability, security cooperation, and substitutive capacity. Reporting should distinguish facts, estimates, management judgments, and unresolved limitations.

7.2. Crosswalk to GRI Reporting Architecture

PAAS adopts GRI 1’s principles of accuracy, balance, clarity, comparability, completeness, sustainability context, timeliness, and verifiability [12]. It aligns governance responsibility and assurance with GRI 2 and material topic processes with GRI 3 [33,34]. The crosswalk does not assert that GRI presently mandates algorithmic disclosures. It shows how a public organization using GRI architecture may locate PAAS content without creating a parallel report. Table 6 crosswalks the ten PAAS fields to GRI 1, GRI 2, and GRI 3 and identifies the corresponding assurance evidence.

Table 6. PAAS fields and crosswalk to GRI 1, GRI 2, and GRI 3.

PAAS Field	Required Content	GRI Alignment	Assurance Evidence
1. Identity and purpose	Name, version, owner, service, intended public value.	GRI 1 clarity/completeness; GRI 2-1, 2-6.	Inventory, charter, approval record.
2. Scope, materiality, tier	Decision influence, population, volume, tier and rationale.	GRI 3-1 to 3-3; GRI 1 sustainability context.	Materiality record, tier decision aid, impact assessment.
3. Legal authority	Statute, regulation, delegation, policy, prohibited uses.	GRI 2-23, 2-27.	Legal opinion, authorization, compliance register.
4. Data provenance	Sources, transformations, quality, lawful use, retention, limitations.	GRI 1 accuracy/verifiability; GRI 3-3.	Lineage map, data quality tests, retention record.
5. Model and decision logging	Version, changes, case-level traceability, retention.	GRI 1 timeliness/verifiability.	Change log, decision record, access log.
6. Roles and human review	Governance body, delegated officials, reviewers, separation of duties.	GRI 2-9, 2-12 to 2-14.	Role matrix, approvals, override records.
7. Performance, impacts, resources	Errors, distributional effects, service outcomes, compute/energy or other material resources.	GRI 3-3; GRI 1 balance/comparability.	Validation, monitoring, resource records, methodology.
8. Contestability and remedies	Notice, explanation, appeals, response time, outcomes, correction.	GRI 2-25, 2-26.	Notice, appeal register, remedy and closure evidence.
9. Audit and corrective action	Assurance scope, findings, incidents, overdue actions, management response.	GRI 2-5, 2-27; GRI 1 verifiability.	Audit report, incident register, corrective action tracker.
10. Vendors and substitution	Provider roles, audit rights, changes, portability, exit and fallback status.	GRI 2-6, 2-23, 3-3.	Contract clauses, portability test, recovery exercise.

7.3. Worked Application: CivicBenefits-Eligibility v3.2

Table 7 applies the framework to a realistic but entirely hypothetical automated benefits eligibility system. The institution, system name, volumes, rates, dates, and findings are fictional and illustrate disclosure form rather than empirical performance. The system receives Tier II because it materially affects access to an essential public benefit. It does not meet Tier III triggers because it operates within one program, does not alter an oversight or constitutional function, and has a tested lawful fallback. If the system were expanded nationwide across multiple essential benefits, used for mass fraud sanctions, or made dependent on a non-substitutable vendor, reassessment could produce Tier III.

The worked application reveals two boundary questions. First, a human final signature does not lower the tier when the system materially structures evidence or options. Second, favorable aggregate performance does not remove the need for individual contestability. It also demonstrates how PAAS makes limitations reportable: a mismatch rate, appeal outcome, overdue finding, or untested fallback is disclosed as a governance fact rather than concealed in a technical annex.

Table 7. Illustrative completed PAAS for CivicBenefits-Eligibility v3.2 (all operational details are fictional).

PAAS Field	Completed Illustrative Disclosure
1. Identity and purpose	CivicBenefits-Eligibility v3.2, owned by the Department of Family Support. It verifies application completeness, matches authorized income records, calculates an eligibility recommendation, and presents reasons to a benefits officer. Purpose: reduce processing delay while preserving lawful individualized review.
2. Scope, materiality, tier	Tier II. The system materially influences eligibility and payment timing for a food support benefit. Reporting period: 1 July 2025–30 June 2026. It processed 48,200 applications; 31,740 received a positive recommendation, 12,860 required additional evidence, and 3600 received a negative recommendation before human review.
3. Legal authority	Operates under the fictional Public Nutrition Assistance Act §§12–18 and implementing rule 4.2. Final authority remains with designated benefits officers. Prohibited uses: immigration enforcement, unrelated fraud profiling, and reuse of application data for model training without separate authorization.
4. Data provenance	Inputs: applicant-submitted household data; authorized wage records; program history; current statutory thresholds. Income values are not averaged across unmatched periods. Each source and transformation is versioned. Quarterly quality checks identified a 1.1% wage-record mismatch rate; mismatches require applicant notice and officer review.
5. Model and decision logging	Rules engine v3.2.4 and document classifier v2.1. Every case record stores source timestamps, rule/model version, material inputs, output, uncertainty flag, officer action, override reason, and final disposition for the legally approved retention period. Material changes require new validation and PAAS update.
6. Roles and human review	Program Director: accountable owner; General Counsel delegate: legal approval; Chief Data Officer: provenance; Benefits Operations Manager: use and training; independent Quality Office: validation; named benefits officer: final determination. Negative recommendations and uncertainty flags require human review. Supervisors sample 5% of positive recommendations monthly.
7. Performance, impacts, resources	Illustrative validation: 96.8% rule-consistency rate; document classifier false-negative rate 2.4%; no subgroup difference above the approved 3-percentage-point investigation threshold after adjustment for sample size. Median processing time fell from 12 to 7 days. Resource reporting includes cloud-compute hours and electricity estimate with method and uncertainty; no environmental claim is made without verified data.
8. Contestability and remedies	Every adverse notice states the material facts, rules applied, data sources, officer identity, how to obtain the case record, and a 30-day reconsideration route. A separate officer reviews challenges. During the period, 612 decisions were challenged; 146 were revised after new or corrected evidence. No benefit suspension occurs solely because an automated match is disputed.
9. Audit and corrective action	Independent pre-release validation completed 15 June 2025. Quarterly reviews examine error, overrides, complaints, security, and drift. One medium finding concerned delayed update of a wage-source schema; corrected 18 February 2026 and retested. No critical incidents. Next external assurance review is due 30 September 2026.
10. Vendors and substitutive capacity	Vendor hosts the document classifier under clauses requiring audit cooperation, incident notice within 24 h, version change approval, data return, and export in documented formats. Department retains rules, case records, and manual procedures. A May 2026 exercise restored manual intake within 6 h and cleared the simulated backlog within 3 business days.

7.4. Audit and Assurance Cycle

The assurance cycle has seven stages: inventory and scope; tier assignment; authorization and control design; pre-deployment validation; operation and evidence capture; independent post-deployment audit; and public reporting, corrective action, and reassessment. Material change, incident, complaint trend, legal amendment, new population, new vendor, model replacement, or scope expansion triggers an interim review. Audit criteria should identify the applicable tier, legal obligations, internal controls, and PAAS claims. Findings should state evidence, condition, criterion, consequence, owner, due date, and closure test [35,49].

Assurance independence is functional. A Tier I review may be performed by a competent unit outside daily operation. Tier II requires validation and periodic review free from program delivery incentives. Tier III requires external or statutorily independent access, subject to lawful security protections. Public summaries may protect personal, classified, or security-sensitive information, but withheld detail should be available to an authorized oversight forum and the basis for withholding should be recorded.

7.5. Administrative Cost and Feasibility

Implementation costs arise from inventory creation, tier assignment, documentation, validation, assurance, public reporting, remediation, training, records management, procurement, and continuity testing. The principal efficiency mechanism is the reuse of evidence collected once across management review, impact assessment, audit, appeal, procurement oversight, and public reporting. Table 8 identifies the principal cost drivers and the feasibility measures available to institutions without implying a monetary estimate.

Costs have different profiles. Inventory design, templates, role definitions, training, and procurement clauses are largely fixed institutional costs. Decision logging, notice, case review, and monitoring vary with system volume and complexity. Independent validation, accessibility work, and continuity exercises recur by tier and change cadence. Incidents, appeals, and remediation are contingent costs. Vendor opacity and fragmented records increase all categories because reviewers must reconstruct evidence after the fact.

Small institutions face high fixed costs and limited specialist capacity. Proportionality should reduce duplication, not remove rights protections. A small municipality or agency may use a central government inventory service, shared legal and audit expertise, common templates, framework contracts, pooled testing, and risk-based audit cadence. It may publish one consolidated PAAS covering a common platform while still identifying local authority, decision volume, appeals, incidents, and responsible officials. Higher-level government or a professional body may maintain shared assurance infrastructure. None of these measures justifies eliminating notice, meaningful review, or correction for a person affected by a consequential decision.

Feasibility should be evaluated during pilots through staff time, documentation latency, audit cost, evidence reuse, appeal processing, remediation closure, and service delay. No credible cost estimate follows from the present conceptual analysis. Reporting unverified dollar or person-hour values would create false precision. The proposed cost model instead identifies variables that pilot institutions should measure.

Table 8. Cost drivers and feasibility measures.

Cost Category	Primary Driver	Feasibility Measure	Non-Negotiable Safeguard
Inventory and tiering	Number of systems, unclear ownership, scope drift.	Central register, common taxonomy, annual attestation, decision aid.	Every materially influential system receives a recorded classification.
Documentation and logs	Legacy systems, fragmented data, case volume.	Standard evidence schema, automated capture, proportionate retention, shared records services.	Enough evidence remains to reconstruct a contested decision.
Validation and audit	Tier, novelty, affected population, change frequency.	Pooled assurance teams, shared-service auditors, accredited external panels, common test suites.	Independence and competence match consequence.
Notice and appeals	Adverse-decision volume, accessibility needs, complexity.	Reusable notice components, case-record portal, triage, trained independent reviewers.	Meaningful human reconsideration and remedy remain available.
Reporting	Multiple frameworks and reporting calendars.	Embed PAAS in GRI/integrated/governance report; reuse impact and audit evidence.	Balanced disclosure includes limitations, incidents, and overdue actions.
Vendor and continuity	Proprietary systems, concentration, weak portability.	Model clauses, documentation escrow where lawful, standard exports, joint recovery exercises.	Public authority retains explanation, audit access, and tested fallback.

8. Discussion

8.1. Interpretation of the Design Propositions

DP1 is articulated and instantiated as design logic within the artifact. The cases and derivation trail show why system-level documentation alone does not answer who possessed authority, who accepted responsibility, how an individual decision was formed, or how a forum reviewed it. The public decision configuration therefore serves as the framework's audit object. This articulation does not establish an empirical effect. Future studies should compare reconstruction time, evidentiary completeness, and reviewer agreement under system-only and decision-configuration approaches.

DP2 is instantiated through the tier triggers and cumulative requirements. Differentiated obligations in the EU, South Korea, Canada, and U.S. federal guidance provide comparative design grounding [8,29–32,36,37], while the proposed tiers add a distinct public-authority axis: consequences for rights, civil service, essential services, institutional independence, and continuity. Neither the comparative presence of tiering nor the worked example constitutes empirical confirmation. Pilot testing should measure inter-rater agreement on tier assignments and refine thresholds.

DP3 is instantiated through the PAAS–GRI crosswalk. PAAS maps decision-level content to established GRI principles and governance disclosures [12,33,34], which articulates design coherence for completeness, comparability, and assurance. The crosswalk does not demonstrate that readers understand or use the disclosures. Experimental and field research should compare PAAS with inventories, impact assessments, and narrative technology reports.

DP4 is instantiated through the shared evidence schema. Authorization records, lineage, logs, impact assessments, appeals, audit findings, and recovery tests serve multiple governance processes. This design logic identifies reuse opportunities without demonstrating savings. Actual effects depend on system maturity, legacy architecture, volume, and assurance design. Pilot studies should quantify evidence reuse and identify whether documentation becomes defensive, duplicative, or disconnected from decisions.

8.2. Theoretical and Standards Contribution

The framework extends public accountability theory by specifying the evidence through which the actor–forum relationship operates in an algorithmically mediated decision [17]. It extends algorithmic auditing by placing lifecycle controls inside legal authority, administrative review, and public reporting [14–16]. It extends hybrid intelligence governance by moving from human–AI workflow controls to institution-level authorization, external assurance, and governance disclosure [50].

The standards contribution is deliberately modular. ISO/IEC 42001 or NIST AI RMF may govern internal management; an AIA may assign impact; ISO 19011 may guide audit; procurement rules may govern vendors; and GRI may structure public reporting. PAAS does not replace those instruments. It supplies the missing join among them: one set of decision-level facts and evidence references, escalated according to public authority and disclosed with material limitations and remedies.

8.3. Transferability and Institutional Conditions

Transferability depends on functional equivalence. Legal authority may arise from statute, regulation, ministerial delegation, ordinance, or other lawful instrument. Independent review may be judicial, parliamentary, audit-based, regulatory, ombudsman-based, or administrative. Records, privacy, freedom-of-information, labor, procurement, and security law will alter implementation. The framework should therefore be adopted through a jurisdiction-specific crosswalk that identifies the competent authorizer, review forum, records custodian, assurance provider, reporting vehicle, and remedies.

Political and organizational conditions remain decisive. Standards do not substitute for professional civil service, legislative scrutiny, independent audit, courts or tribunals, journalists, and civil society. Where oversight is intentionally disabled, documentation may become performative. Tier III's independent access, cessation authority, protected records, and substitutive capacity are designed to make that weakness visible, but a standards document does not create institutional independence by itself.

8.4. Limitations and Validation Agenda

The framework has five limitations. First, the research is conceptual and uses one researcher's qualitative coding. Second, formal instruments do not establish consistent implementation. Third, the comparator set is purposive and does not represent all legal traditions or lower levels of government. Fourth, the tier thresholds are normative and include judgment-laden concepts such as material influence and institutional independence. Fifth, the worked PAAS uses fictional data and demonstrates completeness rather than organizational performance.

Validation should proceed in stages. A preregistered multi-rater study should code common case files and estimate agreement on the ten criteria, five primitives, and tier assignment. Pilots in benefits, licensing, procurement, workforce, and service-triage settings should measure documentation time, assurance cost, evidence reuse, error detection, appeal outcomes, remediation closure, and reporting comprehension. Comparative legal studies should test the functional translation of Tier III in parliamentary, federal, local-government,

and limited judicial review systems. Stress tests should examine legacy data, vendor failure, emergency operation, model drift, privacy risk from logs, and attempted scope expansion.

Future research should also evaluate unintended effects: defensive documentation, excessive risk aversion, delayed service, overcollection of personal data, strategic underclassification, and assurance capture by vendors. Empirical calibration should not weaken the universal floor. Its purpose is to improve consistency, proportionality, and administrative usability.

9. Conclusions

Public sector AI governance requires standards that reach beyond model performance and organizational risk management. When computational systems materially influence decisions made under public authority, the institution must demonstrate lawful purpose, provenance, reconstructable decisions, attributable responsibility, effective contestability, independent audit, corrective action, and transparent reporting.

This study used a replicable qualitative protocol, two complementary negative cases, and six mechanism-diverse reference jurisdictions to derive five accountability primitives. It organized those primitives into minimum, heightened, and systemic/constitutional tiers; specified a decision aid and substitutive capacity test; and developed PAAS as a ten-field governance disclosure crosswalked to GRI architecture. A completed hypothetical benefits application demonstrates tier assignment and reporting, while the cost model identifies fixed, variable, recurring, and incident-driven burdens and feasible shared-service responses for small institutions.

The central contribution is not a new list of ethical controls. It is an institutional standards architecture in which technical evidence, public authority, administrative justice, assurance, continuity, and governance reporting operate as one auditable chain. The DOGE–Treasury episode shows why access and authorization precede algorithmic accountability; Robodebt shows the consequences when data, law, notice, review, and responsibility fragment; and the comparators show that legality, testing, impact assessment, differentiated duties, procurement, and vendor controls offer complementary design mechanisms.

The framework now requires empirical testing. Its value will depend on whether independent reviewers assign tiers consistently, institutions reuse evidence rather than multiply paperwork, affected persons obtain more meaningful review, oversight bodies identify problems earlier, and public reports disclose limitations as clearly as claimed benefits. Algorithmic modernization and democratic administration are compatible only when computational capacity remains subordinate to lawful, reviewable, and sustainable public authority.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/standards6030033/s1>, Table S1, Operational Codebook; Table S2, Case $\times$ criterion score matrix; Table S3, Criterion-level evidence, rationale, and limitation notes. All documentary sources cited in the supplementary tables are included in the main-text reference list [21–25,29–32,36,37,39–48].

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The complete coding data underlying the comparative analysis are reported in Supplementary Materials S1, including the operational codebook, case $\times$ criterion score matrix, and criterion-level documentary anchors and rationales. The legal, policy, standards, audit,

ombudsman, and commission documents are publicly available through the sources identified in the reference list. No personal or confidential datasets were created or analyzed.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Alibašić, H. *Hybrid Intelligence for Effective Digital Governance: AI in Administration*; Public Administration and Information Technology; Springer Nature: Cham, Switzerland, 2026; Volume 42, ISBN 978-3-032-28086-2.
2. Criado, J.I.; Sandoval-Almazán, R.; Gil-Garcia, J.R. Artificial Intelligence and Public Administration: Understanding Actors, Governance, and Policy from Micro, Meso, and Macro Perspectives. *Public Policy Adm.* **2025**, *40*, 234–251. [CrossRef]
3. Dunleavy, P.; Margetts, H. Data Science, Artificial Intelligence and the Third Wave of Digital Era Governance. *Public Policy Adm.* **2025**, *40*, 185–214. [CrossRef]
4. ISO/IEC 42001:2023; Information Technology—Artificial Intelligence—Management System. International Organization for Standardization: Geneva, Switzerland; International Electrotechnical Commission: Geneva, Switzerland, 2023.
5. ISO/IEC 23894:2023; Information Technology—Artificial Intelligence—Guidance on Risk Management. International Organization for Standardization: Geneva, Switzerland; International Electrotechnical Commission: Geneva, Switzerland, 2023.
6. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*; NIST AI 100-1; U.S. Department of Commerce: Gaithersburg, MD, USA, 2023. [CrossRef]
7. IEEE 7000-2021; IEEE Standard Model Process for Addressing Ethical Concerns during System Design. Institute of Electrical and Electronics Engineers: New York, NY, USA, 2021. [CrossRef]
8. European Parliament; Council of the European Union. Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, L 2024/1689, 12 July 2024. Available online: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng> (accessed on 13 July 2026).
9. Alibašić, H. *Sustainability and Resilience Planning for Local Governments: The Quadruple Bottom Line Strategy*; Springer: Cham, Switzerland, 2018.
10. Alibašić, H. Sustainability Is Dead: Adaptive Transformation Governance and the Quadruple Bottom Line as a Post-SDG Framework. *Sustain. Dev.* **2026**, 1–10. [CrossRef]
11. Glavič, P. Standards on Corporate and Public Sustainability Reporting. *Standards* **2026**, *6*, 8. [CrossRef]
12. Global Reporting Initiative. *GRI 1: Foundation 2021*; GRI: Amsterdam, The Netherlands, 2021.
13. International Federation of Accountants; Chartered Institute of Public Finance and Accountancy. *International Framework: Good Governance in the Public Sector*; IFAC: New York, NY, USA; CIPFA: London, UK, 2014.
14. Wieringa, M. What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 1–18. [CrossRef]
15. Raji, I.D.; Smart, A.; White, R.N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 33–44. [CrossRef]
16. Ada Lovelace Institute; AI Now Institute; Open Government Partnership. *Algorithmic Accountability for the Public Sector*; Ada Lovelace Institute: London, UK, 2021.
17. Bovens, M. Analysing and Assessing Accountability: A Conceptual Framework. *Eur. Law J.* **2007**, *13*, 447–468. [CrossRef]
18. Busuioc, M. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Adm. Rev.* **2021**, *81*, 825–836. [CrossRef] [PubMed]
19. Novelli, C.; Taddeo, M.; Floridi, L. Accountability in Artificial Intelligence: What It Is and How It Works. *AI Soc.* **2024**, *39*, 1871–1882. [CrossRef]
20. Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; Lennon, M. Explaining the Principles to Practices Gap in AI. *IEEE Technol. Soc. Mag.* **2021**, *40*, 81–94. [CrossRef]
21. The White House. Executive Order 14158: Establishing and Implementing the President’s “Department of Government Efficiency”. 20 January 2025. Available online: <https://www.whitehouse.gov/presidential-actions/2025/01/establishing-and-implementing-the-presidents-department-of-government-efficiency/> (accessed on 27 July 2026).
22. Benson, P.J.; Linebaugh, C.D. *Privacy Act Lawsuits and the Department of Government Efficiency (DOGE)*; CRS Legal Sidebar LSB11370; Congressional Research Service: Washington, DC, USA, 2025.
23. U.S. Government Accountability Office. *Department of Government Efficiency: Treasury Needs to Fully Implement Data Protection Controls*; GAO-26-108131; GAO: Washington, DC, USA, 2026. Available online: <https://www.gao.gov/products/gao-26-108131> (accessed on 27 July 2026).

24. Royal Commission into the Robodebt Scheme; Report; Commonwealth of Australia: Canberra, Australia, 7 July 2023. Available online: <https://robodebt.royalcommission.gov.au/publications/report> (accessed on 27 July 2026).
25. Commonwealth Ombudsman. *Centrelink's Automated Debt Raising and Recovery System*; Report No. 02/2017; Commonwealth Ombudsman: Canberra, Australia, 2017. Available online: https://www.ombudsman.gov.au/_data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf (accessed on 27 July 2026).
26. Seawright, J.; Gerring, J. Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Polit. Res. Q.* **2008**, *61*, 294–308. [CrossRef]
27. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A Design Science Research Methodology for Information Systems Research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [CrossRef]
28. Gregor, S.; Hevner, A.R. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Q.* **2013**, *37*, 337–355. [CrossRef]
29. Treasury Board of Canada Secretariat. *Directive on Automated Decision-Making*; Government of Canada: Ottawa, ON, Canada, 2025. Available online: <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> (accessed on 27 July 2026).
30. Treasury Board of Canada Secretariat. *Algorithmic Impact Assessment Tool*; Government of Canada: Ottawa, ON, Canada, 2026. Available online: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html> (accessed on 27 July 2026).
31. Executive Office of the President, Office of Management and Budget. *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*; Memorandum M-25-21; Washington, DC, USA, 2025. Available online: <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf> (accessed on 27 July 2026).
32. Executive Office of the President, Office of Management and Budget. *Driving Efficient Acquisition of Artificial Intelligence in Government*; Memorandum M-25-22; Washington, DC, USA, 2025. Available online: <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf> (accessed on 27 July 2026).
33. Global Reporting Initiative. *GRI 2: General Disclosures 2021*; GRI: Amsterdam, The Netherlands, 2021. Available online: <https://www.globalreporting.org/pdf.ashx?id=12358> (accessed on 27 July 2026).
34. Global Reporting Initiative. *GRI 3: Material Topics 2021*; GRI: Amsterdam, The Netherlands, 2021. Available online: <https://www.globalreporting.org/pdf.ashx?id=12453> (accessed on 27 July 2026).
35. ISO 19011:2018; Guidelines for Auditing Management Systems. International Organization for Standardization: Geneva, Switzerland, 2018.
36. Republic of Korea. Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation for Trust, Act No. 20676; Enacted 21 January 2025; Effective 22 January 2026. Available online: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=71019&lang=ENG (accessed on 13 July 2026).
37. Library of Congress. *South Korea: Comprehensive AI Legal Framework Takes Effect*; Global Legal Monitor: Washington, DC, USA, 2026. Available online: <https://www.loc.gov/item/global-legal-monitor/2026-02-20/south-korea-comprehensive-ai-legal-framework-takes-effect/> (accessed on 27 July 2026).
38. Organisation for Economic Co-operation and Development. *OECD AI Principles*; OECD: Paris, France, 2019. Available online: <https://oecd.ai/en/ai-principles> (accessed on 27 July 2026).
39. Personal Data Protection Commission Singapore. *Model Artificial Intelligence Governance Framework*, 2nd ed.; PDPC: Singapore, 2020. Available online: <https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework> (accessed on 27 July 2026).
40. Cabinet Office, Government of Japan. *Social Principles of Human-Centric AI*; Cabinet Office: Tokyo, Japan, 2019.
41. Pilving, I.; Mikiver, M. A Kratt as an Administrative Body: Algorithmic Decisions and Principles of Administrative Law. *Jurid. Int.* **2020**, *29*, 47–58. [CrossRef]
42. *Administrative Procedure Act*; Riigi Teataja: Tallinn, Estonia, 2001. Available online: <https://www.riigiteataja.ee/en/eli/527032019002/consolide> (accessed on 27 July 2026).
43. *Taxation Act*; Riigi Teataja: Tallinn, Estonia, 2002. Available online: <https://www.riigiteataja.ee/en/eli/525032019001/consolide> (accessed on 27 July 2026).
44. Infocomm Media Development Authority. *AI Verify—An AI Governance Testing Framework and Toolkit*; IMDA: Singapore, 2026. Available online: <https://www.imda.gov.sg/how-we-can-help/ai-verify> (accessed on 27 July 2026).
45. Allen, J.G.; Loo, J.; Luna, J. Governing Intelligence: Singapore's Evolving AI Governance Framework. *Camb. Forum AI Law Gov.* **2025**, *1*, e12. [CrossRef]
46. Alibašić, H. Harmonizing Artificial Intelligence Governance: A Comparative Analysis of Singapore and France's AI Policies and the Influence of International Organizations. *Glob. Public Policy Gov.* **2025**, *5*, 93–113. [CrossRef]

47. Japan. Act on the Promotion of Research and Development and the Utilization of AI-Related Technologies; Act No. 53 of 2025; Promulgated 4 June 2025; Effective 1 September 2025. Available online: <https://laws.e-gov.go.jp/law/507AC0000000053> (accessed on 13 July 2026).
48. Digital Agency, Government of Japan. *The Guideline for Japanese Government's Procurements and Utilizations of Generative AI for the Sake of Evolution and Innovation of Public Administration*; Version 2.0; Digital Agency: Tokyo, Japan, 2026. Available online: <https://www.digital.go.jp/en/news/decb64eb-f26e-41cb-8d37-f3dd173108b8> (accessed on 27 July 2026).
49. Power, M. *The Audit Society: Rituals of Verification*; Oxford University Press: Oxford, UK, 1997.
50. Alibašić, H. Toward a Standards Framework for Hybrid Intelligence Governance: Integrating Human Judgment and AI Decision Support. *Standards* **2026**, *6*, 20. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.