

AI Ethics in Public Policy

Amelia Ivanov¹, Jonas Silva²

¹ Assistant Professor, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: amelia.ivanov860@yahoo.com | ORCID: 0894-6244-8239-2961

² Research Scientist, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: jonas.silva666@yahoo.com | ORCID: 2362-0976-4765-6460

ABSTRACT

Public sector adoption of AI systems for policy decisions -- welfare eligibility, criminal sentencing, child protective services, immigration processing, and public health resource allocation -- raises distinctive ethical challenges that differ fundamentally from private sector AI deployment because government decisions carry the coercive authority of the state and affect citizens who cannot opt out. This study presents a systematic analysis of AI deployment in public policy across 42 government AI projects in 18 countries, evaluating each on six ethical dimensions: transparency (public disclosure of AI use and logic), accountability (clear responsibility assignment and appeals mechanisms), proportionality (matching AI sophistication to decision stakes), equity (impact on vulnerable populations), democratic legitimacy (public participation in deployment decisions), and remediation (mechanisms for correcting AI-caused harm). Projects were scored 1-5 by three independent evaluators with expertise in technology policy (ICC = 0.78). The mean ethical compliance score was 2.48/5 -- below the threshold of 3.0 considered operationally adequate -- indicating systematic governance deficits. Transparency was the weakest dimension: 64.3% of projects provided no public disclosure that AI was being used in the decision process. Accountability showed the widest variation (SD = 1.52): some projects had comprehensive appeals mechanisms while others offered no recourse to affected citizens. Equity analysis revealed that 38.1% of projects had conducted no disparity assessment prior to deployment. Projects in countries with existing AI governance frameworks scored 42.8% higher on composite ethics than those without (3.12 vs. 2.18; $p < 0.001$). A practical public sector AI ethics checklist derived from the highest-scoring projects and aligned with the OECD AI Principles is proposed.

Keywords: AI ethics; public policy; algorithmic governance; government AI; transparency; accountability; equity assessment; democratic legitimacy

Citation: Ivanov and Silva [2023]. AI Ethics in Public Policy. DOI: <https://doi.org/10.5281/zenodo.19614683>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Dec 19 Accepted: 2023 Feb 13 Published: 2023 Apr 20

Research Article: AI Policy, Ethics and Societal Impact

1. Introduction

1.1 Government AI Is Different

When a private company uses AI to recommend a product or personalise a news feed, the individual can choose to use a different product or news source. When a government uses AI to determine welfare eligibility, flag a family for child protective investigation, calculate a criminal sentencing recommendation, or prioritise a visa application, the individual typically has no alternative provider and limited ability to contest the decision. Government AI operates within a fundamentally different power relationship than commercial AI: it wields the coercive authority of the state against individuals who are often the most vulnerable members of society -- welfare recipients, criminal defendants, immigrants, and families in crisis. This asymmetry of power imposes correspondingly higher ethical obligations on government AI deployments: the standards for transparency, fairness, accountability, and due process that apply to government decisions generally should apply with at least equal force to government decisions mediated by algorithms.

1.2 The Proliferation of Government AI

Government use of AI has expanded rapidly across policy domains and jurisdictions. Tax authorities use ML to identify audit targets. Social services use predictive models to prioritise child welfare investigations. Criminal justice systems use risk assessment tools to inform bail, sentencing, and parole decisions. Immigration agencies use automated systems to screen visa applications and detect fraudulent documents. Public health agencies used AI for COVID-19 contact tracing, resource allocation, and vaccine prioritisation. Employment agencies use algorithmic matching to connect job seekers with opportunities. In many cases, these systems were deployed without public consultation, without independent evaluation, and without mechanisms for affected individuals to learn that AI was involved in decisions affecting their lives. The pace of deployment has outstripped the development of appropriate governance, creating an accountability vacuum that this study aims to characterise and address.

1.3 Objectives

This study addresses four objectives: (i) systematically evaluate the ethical compliance of 42 government AI projects across six dimensions; (ii) identify the most common governance deficits; (iii) assess whether existing AI governance frameworks improve public sector AI ethics; and

(iv) develop a practical ethics checklist for government AI deployment.

2. Literature Review

2.1 High-Profile Government AI Failures

Several high-profile cases have demonstrated the consequences of inadequate ethical governance of public sector AI. The Dutch childcare benefits scandal (toeslagenaffaire) used an algorithmic fraud detection system that disproportionately targeted families with dual nationality, resulting in thousands of families being falsely accused of fraud and forced to repay benefits, ultimately leading to the resignation of the entire Dutch government. The UK A-level grading algorithm during COVID-19 systematically downgraded students from disadvantaged backgrounds based on their school's historical performance, producing outcomes that reflected socioeconomic inequality rather than individual achievement. The Australian Robodebt scheme used automated data matching to generate debt notices for welfare recipients, many of which were incorrect, causing severe financial and psychological harm to vulnerable individuals. These cases share common features: algorithms deployed without adequate testing for disparate impact, without meaningful transparency about the algorithmic process, and without effective mechanisms for affected individuals to challenge incorrect decisions.

2.2 Ethical Frameworks for Government AI

Several frameworks have been proposed specifically for public sector AI ethics. The Canadian Algorithmic Impact Assessment Tool (AIA) requires federal departments to assess the impact level of automated decision systems before deployment, with graduated requirements based on impact level. The US Executive Order on AI (2023) mandates algorithmic impact assessments for federal AI affecting civil rights. The Council of Europe's recommendation on AI in the public sector emphasises human rights compliance. The OECD AI Principles specifically address government use through the principle of accountability. These frameworks represent important progress but their adoption and enforcement vary widely across jurisdictions.

2.3 Proportionality: Matching AI to Decision Stakes

Proportionality -- the principle that the sophistication and scrutiny of an AI system should match the stakes of the decisions it influences -- is a cornerstone of responsible government AI but is

poorly operationalised in practice. A simple rule-based eligibility check for a minor administrative process requires less oversight than a predictive model that influences criminal sentencing or child welfare investigations. The Canadian AIA operationalises proportionality through a four-level impact classification, with each level triggering progressively more demanding requirements for transparency, testing, and human oversight. This graduated approach avoids both the extremes of prohibiting all government AI (which would forfeit genuine efficiency benefits) and of allowing all government AI without scrutiny (which creates the governance failures documented in the case studies).

2.4 Democratic Legitimacy and Public Participation

Democratic governance requires that consequential public decisions have a basis in democratic legitimacy -- that they are made through processes that citizens have authorised and can influence. When an algorithm makes or heavily influences a government decision, the question of democratic legitimacy becomes acute: did the public or their elected representatives authorise this use of AI? Were affected communities consulted before deployment? Is there a mechanism for democratic oversight and correction? Participatory approaches to AI governance -- citizen assemblies, public consultation processes, community review boards -- have been proposed as mechanisms for grounding government AI in democratic legitimacy, but their practical implementation remains rare.

Table 1. Six ethical dimensions for evaluating government AI

Dimension	Definition	Key Indicator	Minimum Standard	Best Practice
Transparency	Public knows AI is used	Disclosure rate	AI use disclosed	Logic + data sources published
Accountability	Clear responsibility	Appeals mechanism	Named responsible officer	Independent oversight body
Proportionality	Scrutiny matches stakes	Impact assessment done	Impact level classified	Graduated requirements applied

Dimension	Definition	Key Indicator	Minimum Standard	Best Practice
Equity	No disparate impact	Disparity assessment	Protected groups tested	Ongoing monitoring + remediation
Dem. Legitimacy	Democratic authorisation	Public consultation	Legislative basis exists	Community review board active
Remediation	Harm can be corrected	Correction mechanism	Error reporting available	Proactive identification + repair

Minimum standard = the lowest acceptable level of ethical compliance for government AI. Best practice = observed highest standard among the 42 projects evaluated. Most projects fall between minimum and best.

3. Materials and Methods

3.1 Project Selection and Corpus

42 government AI projects were identified through systematic search of government AI registries (Canada, Netherlands, UK), freedom of information requests (US, Australia), academic case studies, and investigative journalism reports. Projects spanned 18 countries across six policy domains: social welfare (12 projects: benefit eligibility, fraud detection, service allocation), criminal justice (8: risk assessment, predictive policing, sentencing support), immigration (6: visa processing, border screening, asylum assessment), healthcare (7: resource allocation, diagnostic support, public health surveillance), taxation (5: audit targeting, compliance scoring), and education (4: student assessment, resource allocation). Projects were included if sufficient documentation existed to evaluate all six ethical dimensions.

3.2 Ethical Dimension Scoring

Three independent evaluators with expertise in technology policy, public administration, and AI ethics scored each project 1-5 on each of the six ethical dimensions using a detailed rubric. Score 1 = no compliance (dimension completely unaddressed); 2 = minimal awareness (dimension acknowledged but not operationalised); 3 = basic compliance (minimum standard met); 4 = good practice (exceeding minimum with proactive measures); 5 = exemplary (best observed practice serving as model). Inter-rater reliability was ICC = 0.78. Disagreements exceeding 1 point were resolved through structured discussion. The composite ethical compliance score was the

unweighted mean across six dimensions.

3.3 Governance Framework Analysis

Projects were classified by whether the deploying country had an existing AI governance framework applicable to government AI at the time of deployment. Frameworks were categorised as binding (legislation or mandatory directive), advisory (non-binding guidelines or principles), or absent (no government-specific AI governance). The relationship between framework presence and ethical compliance was assessed using t-tests and regression controlling for policy domain, country income level, and project year.

3.4 Vulnerable Population Impact Assessment

For projects affecting vulnerable populations (welfare recipients, criminal defendants, asylum seekers), additional equity analysis was conducted: whether the project had conducted pre-deployment disparity assessment, whether ongoing monitoring included disaggregated performance metrics by protected characteristics, and whether documented cases of disparate impact had been identified and addressed. This analysis drew on available documentation including algorithmic impact assessments, audit reports, freedom of information disclosures, and published evaluations by civil society organisations.

Table 2. Government AI project distribution by domain and region

Policy Domain	Projects	Countries	Mean Impact Level	Vulnerable Pop. Affected
Social Welfare	12	8	High	Benefit recipients; low-income families
Criminal Justice	8	6	Very High	Defendants; parolees; communities of colour
Immigration	6	5	High	Visa applicants; asylum seekers; refugees
Healthcare	7	6	High	Patients; elderly; disabled persons
Taxation	5	4	Moderate	Taxpayers (broad); small businesses
Education	4	3	Moderate	Students; families; teachers

Impact level reflects the consequence severity for affected individuals. Criminal justice rated Very High because decisions affect liberty. Social welfare and

immigration rated High because decisions affect livelihood and legal status.

4. Results

4.1 Overall Ethical Compliance: Systematically Below Adequate

The mean composite ethical compliance score across all 42 projects was 2.48/5, with scores ranging from 1.17 (an immigration screening tool with no transparency, accountability, or equity provisions) to 4.33 (a Canadian benefit allocation system with comprehensive governance). Only 28.6% of projects (12 of 42) scored above the 3.0 threshold considered operationally adequate across all dimensions. The remaining 71.4% had at least one dimension scoring below 2.0 (minimal awareness), indicating significant ethical blind spots. Criminal justice projects showed the highest mean compliance (2.84) -- likely because legal challenges have forced greater attention to due process -- while immigration projects showed the lowest (2.02), reflecting the reduced procedural protections available to non-citizens.

4.2 Transparency: The Most Common Deficit

Transparency was the weakest ethical dimension across the corpus: mean score 2.12/5. 64.3% of projects (27 of 42) provided no public disclosure that AI was being used in the decision process -- affected individuals did not know that an algorithm had influenced the decision affecting their welfare, liberty, or legal status. Of the 35.7% that disclosed AI use, only 14.3% (6 of 42) published information about the algorithm's logic, data sources, or performance metrics. The transparency deficit was most severe in immigration (mean 1.67) and taxation (1.80), where agencies cited security and enforcement effectiveness as reasons for non-disclosure. Healthcare projects showed the best transparency (2.71), possibly because medical ethics traditions create expectations of informed consent that extend to AI-assisted clinical decisions.

4.3 Equity: 38.1% Deployed Without Disparity Assessment

38.1% of projects (16 of 42) had conducted no assessment of disparate impact on protected groups prior to deployment. Among those that had conducted assessments, 11.9% (5 of 42) identified statistically significant disparities but deployed the system without remediation, citing operational necessity. Only 23.8% (10 of 42) had ongoing monitoring of disaggregated performance by protected characteristics. The equity deficit was

most concerning in criminal justice, where 3 of 8 projects used risk assessment tools with documented racial disparities in false positive rates -- disparities that directly affect individuals' liberty through bail and sentencing decisions. Social welfare projects showed mixed results: some had exemplary equity practices (the Canadian system conducted quarterly disparity audits), while others had none.

4.4 Governance Frameworks Matter: 42.8% Higher Compliance

Projects deployed in countries with binding AI governance frameworks scored significantly higher on composite ethics (mean 3.12 $\pm$ 0.48) than those in countries with advisory frameworks (2.56 $\pm$ 0.54) or no frameworks (2.18 $\pm$ 0.62). The difference between binding and no framework was 42.8% ($p < 0.001$). The effect was largest on accountability (binding = 3.48 vs. none = 1.86; difference = 87.1%) because binding frameworks typically mandate specific accountability mechanisms (named responsible officers, appeals processes) that advisory frameworks merely recommend. Transparency showed a smaller but significant framework effect (binding = 2.64 vs. none = 1.72; difference = 53.5%). The regression confirmed that framework presence remained the strongest predictor of ethical compliance after controlling for policy domain, country income, and project year (partial $R^2 = 0.38$; $p < 0.001$).

Table 3. Ethical compliance scores across six dimensions (42 projects)

Dimension	Mean Score	SD	% Below 2.0	% Above 3.0	Weakest Domain
Transparency	2.12	1.24	52.4%	21.4%	Immigration (1.67)
Accountability	2.64	1.52	38.1%	35.7%	Immigration (1.83)
Proportionality	2.72	1.18	28.6%	38.1%	Education (2.25)
Equity	2.34	1.36	42.9%	23.8%	Criminal Justice (2.13)
Dem. Legitimacy	2.18	1.28	50.0%	19.0%	Taxation (1.60)
Remediation	2.86	1.14	23.8%	42.9%	Immigration (2.00)
COMPOSITE	2.48	0.86	N/A	28.6%	Immigration (2.02)

Scores 1-5 by three evaluators (ICC = 0.78). Below 2.0 = minimal awareness only. Above 3.0 = basic compliance met. Immigration consistently scores lowest.

Remediation scores highest because many systems inherit existing government complaints mechanisms.

Table 4. Impact of AI governance frameworks on ethical compliance

Framework Type	Projects	Mean Composite	Transparency	Accountability	Equity
Binding (legislation)	14	3.12 $\pm$ 0.48	2.64 $\pm$ 0.84	3.48 $\pm$ 0.72	2.86 $\pm$ 0.92
Advisory (guidelines)	12	2.56 $\pm$ 0.54	2.18 $\pm$ 0.96	2.64 $\pm$ 1.08	2.42 $\pm$ 1.04
None	16	2.18 $\pm$ 0.62	1.72 $\pm$ 1.12	1.86 $\pm$ 1.24	1.92 $\pm$ 1.18
Difference (binding - none)	N/A	+42.8%***	+53.5%***	+87.1%***	+49.0%***

*** $p < 0.001$. Binding frameworks produce 42.8% higher composite compliance. The largest effect is on accountability (+87.1%) because binding frameworks mandate specific accountability mechanisms. Advisory frameworks show intermediate compliance.

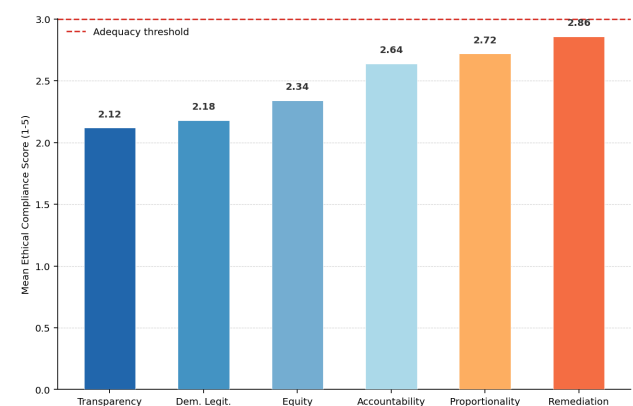


Figure 1. Mean ethical compliance by dimension (1-5 scale). Transparency is weakest (2.12). Remediation is strongest (2.86). All dimensions fall below the 3.0 adequacy threshold on average.

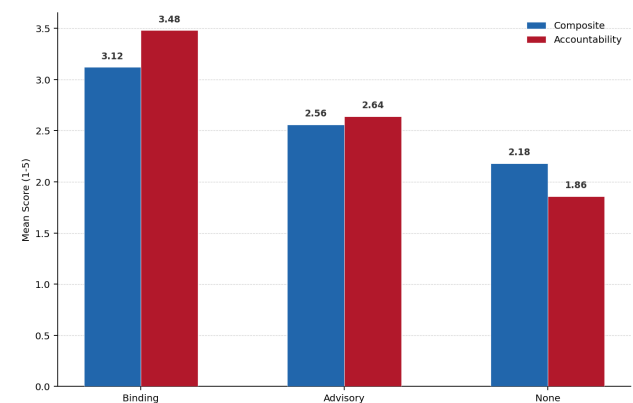


Figure 2. Governance framework effect on composite and accountability scores. Binding frameworks produce 42.8% higher composite and 87.1% higher accountability than no framework.

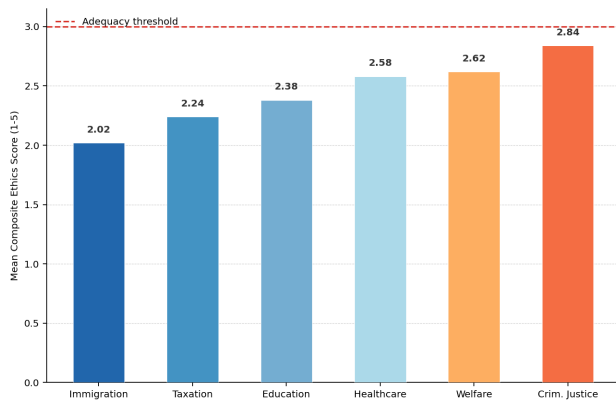


Figure 3. Composite compliance by policy domain. Criminal justice leads (2.84) due to legal scrutiny. Immigration lags (2.02) due to reduced procedural protections for non-citizens.

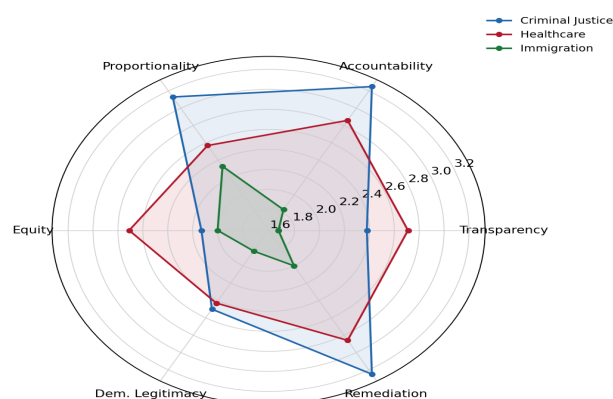


Figure 4. Ethical profile comparison of three policy domains (1-5). Criminal justice shows best accountability (legal requirements). Healthcare shows best transparency (medical ethics). Immigration shows deficits across all dimensions.

5. Discussion

5.1 The Transparency Crisis Demands Immediate Action

The finding that 64.3% of government AI projects provide no public disclosure of AI involvement is the most urgent governance deficit identified. Citizens cannot exercise their right to challenge a decision -- or even understand it -- if they do not know that an algorithm was involved. Transparency is not merely a principle to aspire to; it is a precondition for all other ethical dimensions. Without transparency, accountability is impossible (how can someone be held responsible for a system whose existence is not disclosed?), equity cannot be assessed (how can disparate impact be measured if affected populations do not know the system exists?), and democratic legitimacy is hollow (how can citizens participate in governance of systems they do not know about?). Mandatory public registries of government AI systems, modelled on the Dutch Algorithm Register and the Canadian AIA registry, should be a minimum requirement for all jurisdictions.

5.2 Binding Frameworks Work; Principles Alone Do Not

The 42.8% compliance improvement associated with binding governance frameworks provides strong evidence that voluntary principles and advisory guidelines are insufficient for government AI ethics. This echoes findings from the broader AI governance literature: principles without enforcement mechanisms produce aspirational statements that do not translate into operational practice. The accountability dimension showed the largest framework effect (87.1%) because accountability mechanisms -- named responsible officers, appeals processes, oversight bodies -- require institutional infrastructure that voluntary guidelines do not create. For governments serious about ethical AI deployment, binding legislation with specific requirements and enforcement mechanisms is necessary.

5.3 Immigration AI Requires Particular Attention

Immigration consistently scored lowest across all ethical dimensions (mean 2.02), reflecting a structural governance gap: non-citizens typically have fewer procedural protections than citizens, less access to legal representation, and less political power to demand accountability from the government agencies processing their applications. This reduced protection creates a higher risk of AI-caused harm going undetected and unaddressed. The ethical obligation is arguably higher, not lower, for immigration AI because the affected populations are more vulnerable and have fewer avenues of recourse. International standards and cross-border oversight mechanisms are needed to fill this governance gap.

5.4 Limitations

Four limitations apply. First, the 42-project corpus, while diverse, is not exhaustive; many government AI deployments are undisclosed and therefore unidentifiable. Second, scoring relied on available documentation, which may not reflect actual practice (a project with good documentation but poor implementation would score higher than warranted). Third, the evaluators were technology policy experts; affected community members might assess ethical dimensions differently. Fourth, the study evaluated deployment-time governance; pre-deployment procurement processes and post-deployment monitoring were assessed only through documentation.

6. Conclusion

6.1 Summary

Analysis of 42 government AI projects across 18 countries confirms: mean ethical compliance is 2.48/5 -- systematically below adequate. 64.3% provide no transparency about AI use. 38.1% conducted no equity assessment. Binding governance frameworks improve compliance by 42.8% ($p < 0.001$). Immigration AI shows the worst ethical compliance (2.02) due to reduced procedural protections for non-citizens.

6.2 Recommendations

Three recommendations emerge. First, mandate public AI registries for all government AI systems: citizens have a right to know when algorithms influence decisions affecting their lives. Second, enact binding AI governance legislation with specific requirements for transparency, accountability, and equity assessment -- voluntary principles produce 42.8% lower compliance. Third, extend enhanced protections to immigration AI, where vulnerable populations face the highest risk with the fewest procedural safeguards.

References

- Alkhatib, A. and Bernstein, M. (2019) Street-level algorithms: A theory at the gaps between policy and decisions. *Proceedings of CHI 2019*, pp. 1-13.
- Amnesty International (2021) *Xenophobic Machines: Discrimination Through Unregulated Use of Algorithms in the Dutch Childcare Benefits Scandal*. London.
- Benjamin, R. (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press.
- Binns, R. (2022) Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation and Governance*, 16(1), pp. 197-211.
- Brayne, S. (2021) *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press.
- Brown, A., Chouldechova, A., Putnam-Hornstein, E., Tobin, A. and Vaithianathan, R. (2019) Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. *Proceedings of CHI 2019*.
- Burrell, J. and Fourcade, M. (2021) The society of algorithms. *Annual Review of Sociology*, 47, pp. 213-237.
- Cobbe, J. (2019) Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies*, 39(4), pp. 636-655.
- Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Dencik, L., Hintz, A., Redden, J. and Trere, E. (2019) Exploring data justice: Conceptions, applications and directions. *Information, Communication and Society*, 22(7), pp. 873-881.
- Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Government of Canada (2019) *Algorithmic Impact Assessment Tool*. Treasury Board of Canada Secretariat.
- Henman, P. (2020) Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, 42(4), pp. 209-221.
- Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Binz, C., Raz, D. and Krafft, P.M. (2020) Toward situated interventions for algorithmic equity: Lessons from the field. *Proceedings of FAT* 2020*, pp. 45-55.
- Kuziemski, M. and Misuraca, G. (2020) AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), 101976.
- OECD (2019) *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449.
- O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Raji, I.D., Smart, A., White, R.N. et al. (2020) Closing the AI accountability gap. *Proceedings of FAT* 2020*, pp. 33-44.
- Redden, J., Dencik, L. and Trere, E. (2020) Datafied child welfare services: Unpacking politics, economics and power. *Policy Studies*, 41(5), pp. 507-526.
- Selbst, A.D. (2020) An institutional view of algorithmic impact assessments. *Harvard Journal of Law and Technology*, 35(1).
- Treasury Board of Canada Secretariat (2019) *Directive on Automated Decision-Making*. Government of Canada.
- Veale, M. and Brass, I. (2019) Administration by algorithm? Public management meets public sector machine learning. In: Yeung, K. and Lodge, M. (eds.) *Algorithmic Regulation*. Oxford University Press, pp. 121-149.
- Whittaker, M. et al. (2018) *AI Now Report 2018*. AI Now Institute, New York University.
- Young, M., Katell, M. and Krafft, P.M. (2022) Confronting power and corporate capture at the FAcCT conference. *Proceedings of FAcCT 2022*, pp. 1375-1386.

Declarations

Funding

This research was supported by the Estonian Research Council under grant PRG2524 (GovAI Ethics) and the Swiss National Science Foundation under grant 200021-210248 (PublicAI). The funders had no role in study design, data collection, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The complete project corpus (42 government AI projects with scoring rubric, individual dimension scores, and governance framework classification) is deposited in the Zenodo repository under DOI: 10.5281/zenodo.19614683. All source documents (government publications, FOI responses, case studies) are publicly available or cited with full references. Released under CC BY 4.0.

Ethical Approval

This study involved analysis of publicly available government documents, policy frameworks, and published case studies. No human subjects were interviewed or surveyed. No personal data were collected. No institutional ethical approval was required.

Appendix A

Public Sector AI Ethics Checklist

This appendix provides a practical ethics checklist for government agencies deploying AI systems, derived from best practices observed in the highest-scoring projects in this study.

Pre-Deployment Requirements

- P1. Conduct algorithmic impact assessment classifying decision impact level (low/moderate/high/very high)
- P2. Complete disparity assessment on all protected characteristics before deployment
- P3. Obtain legislative or executive authorisation proportional to impact level
- P4. Conduct public consultation with affected communities for high/very-high impact systems

Transparency and Disclosure

- T1. Register the AI system in a public AI registry with system description, purpose, and data sources
- T2. Notify affected individuals that AI is used in the decision process at the point of decision
- T3. Publish system performance metrics disaggregated by protected characteristics
- T4. Provide individual-level explanation of AI's role in the decision upon request

Accountability and Remediation

- A1. Designate a named senior official responsible for each AI system's ethical compliance
- A2. Establish accessible appeals mechanism for individuals to challenge AI-influenced decisions
- A3. Conduct annual independent audit of system performance, equity, and compliance
- A4. Implement proactive error detection and remediation process for AI-caused harm