Artificial Intelligence, Governance and the Public Interest in Digital Societies

Miriam Cole, Elias Courtney

Article type: Editorial

Volume 1, Issue 1 (2026)

Read Current Issue →

Abstract

Artificial intelligence is becoming embedded in institutional decision-making, public services, education, health systems, finance, regulation, and digital platforms. Its public significance therefore depends not only on model capability but on the organizational, legal, and social arrangements through which automated systems are selected, configured, supervised, contested, and evaluated. This editorial argues that governing AI in the public interest is fundamentally a problem of institutional capacity and accountability. Bias, opacity, data quality, automation bias, privacy risks, vendor dependence, weak procurement, and fragmented responsibility can undermine otherwise capable systems, while formal transparency alone may fail when affected people cannot understand or challenge decisions. Effective governance requires independent auditing, meaningful documentation, contestability, human oversight, public-sector expertise, procurement standards, longitudinal evaluation, and evidence about actual social outcomes after deployment. Identical technologies can produce different consequences across organizations because professional norms, legal safeguards, incentives, data infrastructures, and administrative practices shape how AI is used. Research should therefore evaluate AI as a socio-technical institutional system rather than as an isolated technical artefact.

Keywords: AI governance; algorithmic accountability; public-sector AI; automated decision-making; institutional capacity; transparency; public interest

Citation

Cole, M., & Courtney, E. (2026). Artificial Intelligence, Governance and the Public Interest in Digital Societies. Digital Society, Governance & Development, 1(1). https://eldranacademic.com/dsgd/articles/1/

Received	Revised	Accepted	Published
15 December 2025	10 February 2026	26 February 2026	12 March 2026

Copyright

© 2026 Miriam Cole, Elias Courtney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) ↗. Use, distribution or reproduction in other forums is permitted, provided the original authors and copyright owners are credited and the original publication in this journal is cited in accordance with accepted academic practice.

Disclaimer

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, the publisher, the editors or the reviewers. Any product evaluated in this article, or claim made by its manufacturer, is not guaranteed or endorsed by the publisher.

1.0 Introduction

1.1 Context and Problem Framing

Artificial intelligence is no longer confined to experimental laboratories or private technology firms. Automated decision systems increasingly support eligibility screening, fraud detection, risk classification, resource allocation, inspection, recruitment, customer service, educational analytics, clinical decision support, financial services, content moderation, and administrative case management. Public organizations are drawn to these systems because they promise greater speed, consistency, personalization, and capacity to analyse large volumes of data. Wirtz et al. (2019) identify a wide range of public-sector AI applications, while Sun and Medaglia (2019) show that adoption simultaneously creates organizational, data, technical, legal, ethical, and social challenges. More recent evidence indicates that public institutions are beginning to build governance processes around AI rather than treating ethical principles as external guidance alone (de Almeida & dos Santos Júnior, 2024). International policy frameworks likewise increasingly frame public-sector AI as a lifecycle governance problem involving data, people, procurement, deployment, monitoring, and accountability (OECD & UNESCO, 2024; OECD, 2025).

The central tension is that institutional use changes the stakes of technical performance. An inaccurate recommendation in a low-consequence consumer application is different from an automated assessment that influences access to benefits, employment, education, health services, credit, or regulatory attention. Algorithmic systems can reproduce historical inequalities when labels and training data reflect biased institutions, and even apparently neutral variables may function as proxies for protected or disadvantaged characteristics (Barocas & Selbst, 2016). Fairness cannot therefore be resolved solely by optimizing mathematical metrics because social categories, institutional objectives, and decision procedures shape what counts as a fair outcome (Selbst et al., 2019). Opacity creates a related problem: even when code or model documentation is available, organizational complexity may prevent affected people or frontline professionals from understanding how a decision was produced (Ananny & Crawford, 2018). Governance must consequently concern the full decision process, not only the model.

1.2 Why the Issue Matters

The proposition advanced here is that AI governance should be assessed by whether institutions can justify, supervise, correct, and learn from automated decisions in practice. Accountability involves identifying who has authority to approve deployment, who monitors performance, who responds to errors, and how affected people can challenge consequential outcomes. Kroll et al. (2017) demonstrate that accountable algorithmic systems require procedural and technical mechanisms capable of supporting oversight rather than simple disclosure. Public administration scholarship similarly emphasizes that automation redistributes discretion instead of eliminating it, moving choices into model design, procurement, data management, thresholds, and override procedures (Bullock, 2019; Young et al., 2019). Governing AI in the public interest therefore requires institutional capacity to connect technical evidence with rights, public values, professional judgment, and mechanisms of redress.

2.0 Current Landscape

2.1 AI Adoption

The contemporary AI-governance landscape combines principles, risk frameworks, regulatory requirements, organizational controls, auditing practices, and documentation mechanisms. High-level ethical frameworks have converged around ideas such as fairness, transparency, accountability, privacy, safety, and human agency, although their translation into operational practice remains uneven (Floridi et al., 2018; Jobin et al., 2019). Risk-oriented governance frameworks increasingly attempt to make these principles actionable. The NIST AI Risk Management Framework organizes governance around mapping context, measuring risk, managing it, and establishing cross-cutting governance functions (National Institute of Standards and Technology, 2023). The European Union's Artificial Intelligence Act adds legally differentiated obligations according to risk and use context, demonstrating a wider shift from voluntary principles toward enforceable lifecycle requirements (European Union, 2024).

2.2 Algorithmic Decision-Making

Public-sector implementation nevertheless depends on organizational arrangements. Kuziemski and Misuraca (2020) show that automated decision-making can intensify power asymmetries when public institutions lack adequate governance mechanisms. Zuiderwijk et al. (2021) similarly identify benefits and risks that span service quality, efficiency, privacy, explainability, accountability, and organizational readiness. Empirical work across public organizations indicates that training, standards, internal processes, and coordination are important determinants of AI-governance maturity (de Almeida & dos Santos Júnior, 2024). These findings reinforce the view that governance capacity is not supplied automatically by regulation; it must be embedded in budgets, roles, professional expertise, procurement, records management, and operational routines.

2.3 Regulatory Responses

Algorithmic auditing has become one of the most concrete accountability mechanisms. Raji et al. (2020) propose end-to-end internal auditing that examines system development, documentation, testing, and organizational accountability rather than limiting review to model outputs. Model cards and datasheets provide complementary documentation of intended uses, evaluation conditions, dataset provenance, limitations, and known risks (Gebru et al., 2021; Mitchell et al., 2019). Yet documentation is only meaningful when people have authority and expertise to act on it. The increasingly recognized fallacy of equating technical functionality with successful institutional function is particularly relevant here: a system can produce technically valid outputs while failing to accomplish the public purpose for which it was introduced (Raji et al., 2022).

2.4 Public-Interest Governance

Human interaction also forms part of the governance landscape. People may reject accurate systems because they distrust algorithmic judgment, but they may also over-rely on automation when it appears objective or authoritative. Dietvorst et al. (2015) document algorithm aversion after observed mistakes, while Logg et al. (2019) demonstrate algorithm appreciation under other conditions. Lee (2018) shows that perceptions of fairness vary according to the type of decision and whether it is made by humans or algorithms. Governance therefore requires attention to how professionals and citizens interpret automated systems, not simply whether an explanation mechanism exists.

3.0 Critical Challenges

3.1 Accountability

Bias and unequal impact remain central challenges because automated systems learn from institutions and social environments that are themselves uneven. Statistical parity across one dimension cannot guarantee fair treatment when historical labels, sampling practices, measurement choices, and deployment procedures encode structural inequalities. Mehrabi et al. (2021) catalogue multiple sources of bias across data, algorithms, and evaluation, while Selbst et al. (2019) warn that abstraction can separate technical fairness interventions from the institutional context that gives them meaning. Public organizations therefore require both quantitative assessment and domain-specific inquiry into who is affected, which errors are consequential, and whether outcomes are contestable.

Opacity and fragmented accountability create a second challenge. Complex models, vendor contracts, proprietary data pipelines, and outsourced infrastructure may prevent agencies from explaining or independently verifying system behaviour. Formal disclosure can be inadequate when technical information is inaccessible to citizens or when no institution has the capacity to investigate it. Ananny and Crawford (2018) argue that transparency is limited when it focuses on seeing inside systems without addressing power and institutional relations. Busuioc (2021) similarly emphasizes that AI accountability requires identifiable forums, actors, obligations, and consequences. Vendor dependence can weaken this chain when proprietary claims or contractual arrangements prevent independent scrutiny.

3.2 Transparency

Procurement and institutional expertise are therefore governance issues in their own right. Public agencies may acquire AI systems before they possess the technical personnel, data governance, legal expertise, or evaluation procedures needed to manage them. OECD and UNESCO (2024) identify skills, governance, data, infrastructure, and procurement as recurring implementation concerns. Procurement documents that specify functionality but omit audit access, performance monitoring, retraining responsibilities, incident disclosure, or data portability can create long-term dependence. Such arrangements can shift important public decisions toward vendors without a corresponding transfer of public accountability.

3.3 Unequal Impacts

Contestability remains underdeveloped in many automated processes. People affected by consequential decisions need to know that automation was used, understand the relevant basis for a decision, correct inaccurate data, and seek meaningful review. These requirements are distinct from generic explainability. Kroll et al. (2017) emphasize procedural mechanisms for accountability, while the EU AI Act formalizes human oversight, information, and risk-management obligations for defined classes of systems (European Union, 2024). Without workable appeal routes, an apparently transparent system may remain institutionally unaccountable.

3.4 Institutional Capacity

Post-deployment evaluation is another major weakness. Performance can change as populations, policies, data sources, professional practices, or vendor models evolve. Automated systems may also alter human behaviour and organizational routines, producing effects that were absent during pre-deployment testing. Wirtz et al. (2022) argue for risk-oriented AI governance, but institutional evidence is still thin regarding how organizations detect harms, document incidents, measure unequal outcomes, and retire systems that no longer serve their intended purpose. Public trust depends not on promises of responsible AI but on visible evidence that institutions can recognize and correct failures.

4.0 Discussion and Priorities

4.1 Integrative Discussion

AI governance is often described as if better technical design and better institutional governance were separable tasks. In practice they are interdependent. A model's threshold determines how many cases reach human review; staffing levels determine whether that review is meaningful; procurement contracts determine whether retraining or auditing is possible; professional norms influence whether users challenge automated recommendations; and legal rights determine whether affected people can contest an outcome. The same model can therefore produce different consequences across institutions because governance conditions alter how predictions are translated into decisions.

This perspective helps explain why public-interest governance cannot be reduced to a checklist of ethical principles. Principles such as transparency, fairness, privacy, and accountability can conflict or require contextual interpretation. Excessive transparency may expose security-sensitive information or personal data, while insufficient disclosure weakens contestability. Greater automation can improve consistency while simultaneously increasing automation bias or reducing opportunities for professional judgment. Highly explainable systems may still be based on poor data, while technically fair models may be embedded in inequitable policies. Governance must address these interactions rather than optimizing each characteristic independently.

Institutional incentives also shape outcomes. Agencies under pressure to reduce costs may use AI primarily to increase throughput, while organizations emphasizing professional quality may deploy the same technology as decision support. Vendors may optimize for contract metrics that differ from public value. Frontline workers may develop informal workarounds when system outputs conflict with experience. These dynamics illustrate why socio-technical analysis is indispensable: the governance of AI is simultaneously about code, data, organizations, professions, markets, law, and public legitimacy.

Trust should consequently be treated as an outcome of demonstrable institutional competence rather than a communications objective. Citizens do not need unconditional trust in automated systems; they need justified confidence that systems are evaluated, documented, supervised, appealable, and correctable. Meaningful governance therefore connects technical evidence with institutional responsibility and creates mechanisms through which mistakes can produce learning rather than concealment.

Complementary scholarship informing this synthesis includes European Union. (2024); Kaur (2020); National Institute of Standards and Technology. (2023); OECD (2024); OECD. (2025); Zarsky (2016); and Zerilli (2019). These works broaden the evidentiary base considered across the field.

4.2 Research and Practice Priorities

The first priority is independent, context-sensitive auditing. Audits should assess data provenance, performance, subgroup effects, security, human interaction, organizational incentives, and downstream consequences rather than treating fairness metrics as sufficient. Internal auditing frameworks provide an important foundation, but high-consequence systems also require credible external scrutiny and clear access rights for auditors (Raji et al., 2020).

Second, public-sector AI evaluation should extend beyond pre-deployment accuracy. Institutions need longitudinal evidence on error patterns, appeals, overrides, disparate outcomes, workload, service quality, and behavioural adaptation. Evaluations should compare automated systems with realistic human or organizational baselines rather than idealized alternatives. Evidence about actual social outcomes is particularly important where AI changes access to services or redistributes administrative attention.

Third, transparency should be designed for use. Model cards, dataset documentation, procurement records, impact assessments, and public notices should identify intended purpose, limitations, responsible actors, evaluation results, and routes for complaint (Gebru et al., 2021; Mitchell et al., 2019). Technical disclosure should be complemented by explanations tailored to professionals and affected members of the public.

Fourth, contestability requires operational design. Institutions should establish correction procedures for inaccurate data, human review for consequential decisions, time-bounded appeal processes, and records sufficient to reconstruct how an outcome was produced. These mechanisms should be tested with users rather than assumed to work because they exist on paper.

Fifth, institutional capacity must be treated as infrastructure. Public organizations need multidisciplinary teams capable of evaluating data quality, model performance, procurement, security, privacy, law, accessibility, and social impact. Training is particularly important because governance frameworks are ineffective when decision-makers and frontline users do not understand their obligations (de Almeida & dos Santos Júnior, 2024).

Sixth, procurement standards should require auditability, documentation, version control, incident reporting, data portability, performance monitoring, and clear responsibility for updates. Agencies should avoid contracts that make independent verification technically or legally impossible. Finally, replication and comparative institutional research should examine why similar AI systems produce different outcomes across jurisdictions and organizations. Such evidence can identify which governance arrangements actually improve accountability rather than merely formalizing it.

4.3 Implications for the Journal's Scope

Research on AI governance should be judged by the quality of its evidence about both technology and institutions. Strong contributions will connect algorithmic design with public administration, sociology, information systems, law, organizational studies, communication, development, and professional practice. Studies of governance principles are valuable when they specify how principles are operationalized, measured, and enforced; empirical work is especially important when it examines actual deployment rather than hypothetical acceptance alone.

The field also benefits from research that studies failed implementations, contested systems, procurement processes, audit mechanisms, appeals, organizational learning, and post-deployment effects. Such work can reveal institutional conditions that remain invisible in laboratory evaluations. Comparative studies across countries, sectors, and organizational forms are particularly useful because they help distinguish technology-specific effects from consequences produced by governance context.

Submissions addressing automated public services, algorithmic accountability, auditing, responsible procurement, human oversight, generative AI in institutions, data governance, or public trust should therefore make explicit how technological and organizational mechanisms interact. The broader scholarly objective is not to promote or resist AI as a category, but to establish credible evidence about when, where, and under what safeguards automated systems serve legitimate institutional purposes.

4.4 Conclusion

Artificial intelligence increasingly participates in institutional decisions that distribute opportunities, burdens, attention, and resources. Its governance cannot therefore be delegated to technical teams alone or reduced to generic ethics statements. Public-interest outcomes depend on data quality, model behaviour, procurement, professional practice, legal safeguards, organizational incentives, oversight capacity, and the ability of affected people to challenge consequential decisions.

The most important shift is from governing models to governing decision systems. This means documenting who defines objectives, which data are used, how performance is evaluated, when humans can override automation, what vendors are required to disclose, how incidents are investigated, and how systems are changed or withdrawn when evidence deteriorates. It also means measuring actual institutional and social outcomes over time rather than assuming that technical performance translates directly into public value.

Responsible governance should not demand perfect prediction or complete transparency. It should demand proportionate evidence, identifiable responsibility, meaningful contestability, and continuing institutional capacity to learn from failure. The public interest is served when automation remains subordinate to legitimate purposes and when institutions can explain not only what an AI system predicts, but why it is being used, how its consequences are monitored, and who remains accountable for the decisions made around it.

5.0 Declarations

5.1 Author Contributions

Prof. Miriam Cole and Dr. Elias Courtney contributed to the conceptual development, literature synthesis, critical interpretation, drafting, and revision of the editorial.

5.2 Funding

No external funding was received for the preparation of this editorial.

5.3 Competing Interests

The authors declare that they have no conflicts of interest or competing interests relevant to this editorial.

5.4 Data Availability

No original datasets were generated or analysed for this editorial.

5.5 AI Use Disclosure

Generative AI was used to support literature organization and language drafting during preparation of this editorial. The authors remain responsible for independently verifying the accuracy of the arguments, citations, references, and final submitted text in accordance with the publisher's policy.

6.0 References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. The American Review of Public Administration, 49(7), 751–761. https://doi.org/10.1177/0275074019856123

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293

de Almeida, P. G. R., & dos Santos Júnior, C. D. (2024). Artificial intelligence governance: Understanding how public organizations implement it. Government Information Quarterly, 41(4), 102003. https://doi.org/10.1016/j.giq.2024.102003

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3313831.3376219

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705.

Kuziemski, M., & 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. https://doi.org/10.1016/j.telpol.2020.101976

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1). https://doi.org/10.1177/2053951718756684

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115. https://doi.org/10.1145/3457607

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. https://doi.org/10.1145/3287560.3287596

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

OECD, & UNESCO. (2024). G7 toolkit for artificial intelligence in the public sector. OECD Publishing. https://doi.org/10.1787/421c1244-en

OECD. (2025). Governing with artificial intelligence: The state of play and way forward in core government functions. OECD Publishing. https://doi.org/10.1787/795de142-en

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. (2022). The fallacy of AI functionality. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 959–972. https://doi.org/10.1145/3531146.3533158

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33–44. https://doi.org/10.1145/3351095.3372873

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. https://doi.org/10.1145/3287560.3287598

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. International Journal of Public Administration, 42(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. Government Information Quarterly, 39(4), 101685. https://doi.org/10.1016/j.giq.2022.101685

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. Perspectives on Public Management and Governance, 2(4), 301–313. https://doi.org/10.1093/ppmgov/gvz014

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 41(1), 118–132. https://doi.org/10.1177/0162243915605575

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? Philosophy & Technology, 32, 661–683. https://doi.org/10.1007/s13347-018-0330-6

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 38(3), 101577. https://doi.org/10.1016/j.giq.2021.101577