Algorithmic accountability for the public sector

Learning from the first wave of policy implementation

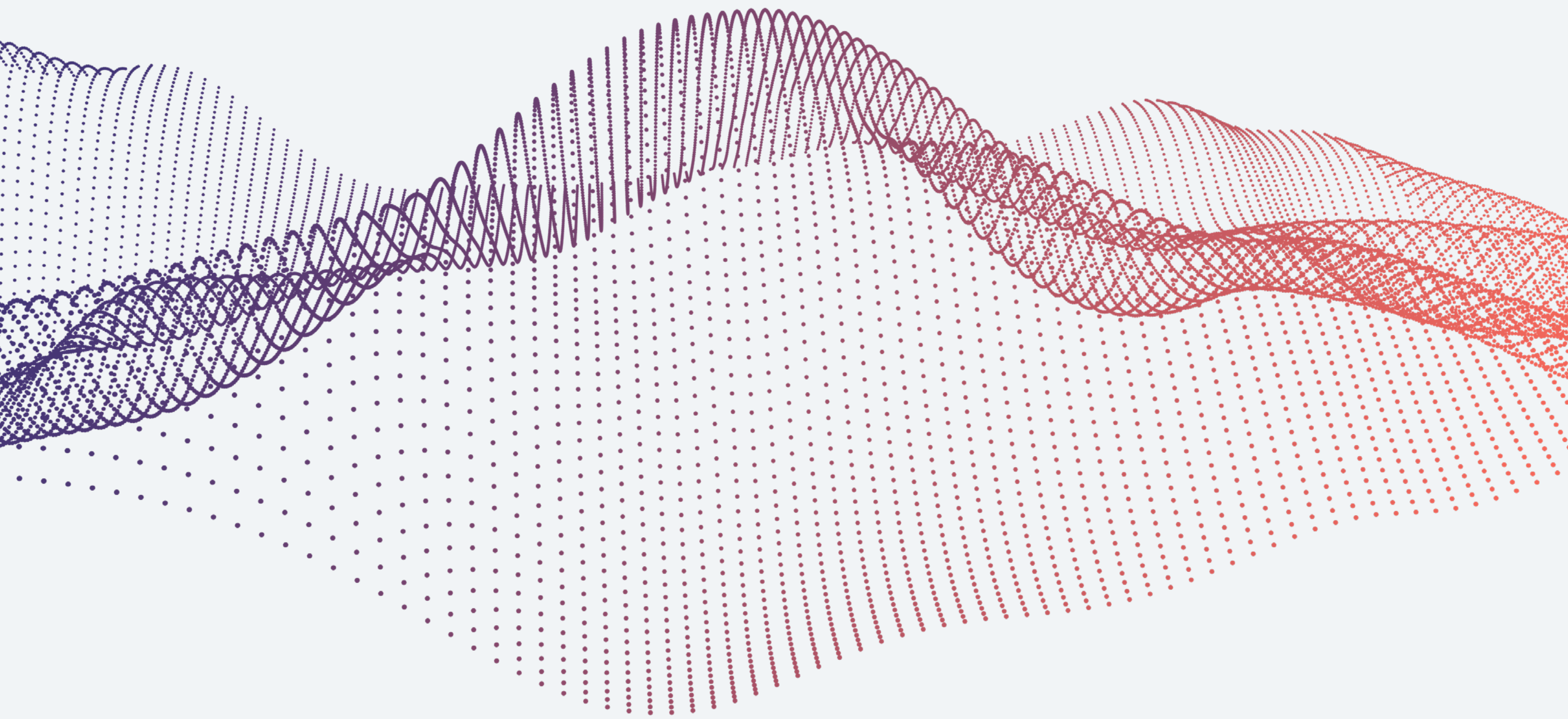
August 2021

EXECUTIVE SUMMARY

The Ada Lovelace Institute (Ada), AI Now Institute (AI Now), and Open Government Partnership (OGP) have partnered to launch the first global study to analyse the initial wave of algorithmic accountability policy for the public sector.

This study aims to understand the challenges and successes of algorithmic accountability policies from the perspectives of the actors and institutions directly responsible for their implementation on the ground. This executive summary highlights the key findings from this study. Read the full report for further detail on these findings and practical case studies of implemented policies.

Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). *Algorithmic Accountability for the Public Sector*. Available at: https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/



Ada Lovelace Institute

AI NOW

Open Government Partnership

Contents

Introduction and scope of study

Governments around the world are increasingly turning to algorithms to automate or support decision-making in public services. Algorithms might be used to assist in urban planning, prioritise social-care cases, make decisions about welfare entitlements, detect unemployment fraud or surveil people in criminal justice and law enforcement settings. The use of algorithms is often seen as a way to improve, increase efficiency or lower costs of public services.

Growing evidence suggests that algorithmic systems in public-service delivery can cause harm and frequently lack transparency in their implementation, including opacity around decisions about whether and why to use them. Most countries have yet to resource efforts to raise awareness and engage the wider public about the use of algorithms in public-service delivery.

In recognition of these conditions, regulators, lawmakers and governmental accountability organisations have turned to regulatory and policy tools, hoping to ensure '**algorithmic accountability**' across countries and contexts. These responses are emergent and fast evolving, and vary widely in form and substance – from legally binding commitments, to high-level principles and guidelines. Lessons from their early implementation raise important challenges and pose questions about the future of governing algorithmic systems.

While there have been some efforts to evaluate algorithmic accountability within particular institutions or contexts,[1] there have been few systematic and cross-jurisdictional studies of the implementation of these policies. This report, commissioned by the Ada Lovelace Institute, AI Now Institute and the Open Government Partnership is the first study to evaluate this initial 'wave' of algorithmic accountability policy for the public sector across jurisdictions. It is focused on the North American and European policy contexts due to the greater number of implemented policies in these regions, and is missing critical perspectives from the Global South. We encourage more research into wider and emerging policy contexts.

This report presents and analyses evidence on the use of algorithmic accountability policies in different contexts from the perspective of those implementing these tools. It also explores the limits of legal and policy mechanisms in ensuring safe and accountable algorithmic systems, and provides practical guidance to the policymakers, civil society, public officials and agencies responsible for implementing related policy tools and commitments. Finally, the report outlines some open questions and generative future directions for the research community in this field.

1 For instance: Young, M., Katell, M., and Krafft, P.M., (2019). 'Municipal surveillance regulation and algorithmic accountability.' *Big Data & Society* 6.2 . Available at: https://journals.sagepub.com/doi/full/10.1177/2053951719868492

Methodology

This report examines the implementation of algorithmic accountability policies by governments in various jurisdictions. The findings of this report are based on:

- A database of more than 40 examples of algorithmic accountability policies at various stages of implementation, taken from more than 20 national and local governments.
- Semi-structured interviews with decision-makers and members of civil society closely involved with the implementation of algorithmic accountability policies in the UK, Netherlands, France, New Zealand, Canada and Chile, as well as at the local level in Amsterdam and New York City.
- Feedback received at a workshop with members of the Informal Network on Open Algorithms[2] that are implementing commitments focusing on algorithmic accountability through their OGP action plans.
- Feedback from participants of a private roundtable at RightsCon 2021 with public officials and members of civil society organisations from many of the countries reviewed in this report.
- A review of existing empirical studies on the implementation of algorithmic accountability policies in various jurisdictions.

The implementation of algorithmic accountability policies varies widely across social, economic, legal and political contexts. Because of this, this study focuses on understanding and analysing the political, institutional and contextual factors that have shaped the implementation of algorithmic accountability policies in various jurisdictions, and how they might enable or perturb the objectives that these policies were purportedly designed to achieve.

Definitions

The term **algorithm** describes a series of steps through which particular inputs can be turned into outputs. An **algorithmic system** is a system that uses one or more algorithms, usually as part of computer software, to produce outputs that can be used for making decisions. We use a functional definition of an algorithmic system, as a system that uses automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans. It is important to note that all algorithmic systems encompass different kinds of human intervention – whether at the stage of design, or in the way they are eventually used. In our analysis, we consider the technical as well as social, cultural, legal and institutional contexts where algorithms are embedded, as important determinants of how these systems are used and governed.

2 Open Government Partnership. *Open Algorithms Network*. Available at: https://www.opengovpartnership.org/about/partnerships-and-coalitions/open-algorithms-network/

In this report, we use the term '**algorithmic accountability policies**' to identify the set of policies oriented towards ensuring that those that build, procure and use algorithms are eventually answerable for their impacts. This terminology builds on the widely used definition of **accountability** provided by Professor Mark Bovens, which describes accountability as a relationship between the *actors* who use or design algorithmic systems, and *forums* that can enforce standards of conduct. This definition of accountability encompasses both the requirement that actors are answerable and can justify their use of algorithmic systems, and also that they can face consequences for such use.[3]

We focus here on accountability mechanisms created or channelled through law and policy. Mechanisms that have emerged to hold algorithmic systems accountable to the contexts and communities they are meant to serve, including tech-worker organising and whistleblowing, community organisers, civil society organisations and investigative journalism, are not examined in this study.

3 Bovens, M., (2007). 'Analysing and assessing accountability: A conceptual framework 1.' *European Law Journal*. Vol. 13. No. 4. pp. 447-468. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0386.2007.00378.x

A typology of algorithmic accountability policies

This section describes the various policy mechanisms through which governments have sought to achieve algorithmic accountability in the public sector, and analyses their theories of change (how they seek to achieve their stated objective). As a relatively recent addition to technological governance, these policies vary widely, as does the vocabulary used to describe them.

The following typology, while not intended to be comprehensive, indicates the forms of algorithmic accountability policies currently taking shape in the public sector.

1. **Principles and guidelines:** A number of policy documents provide non-binding normative guidance, in the form of principles and values, for public agencies to follow. These documents vary in form, but generally identify high-level policy goals, and how they might be implicated by the use of algorithmic systems within public agencies. In some cases, as in the UK Data Ethics Framework,[4] or the Australian Ombudsman's Better Practice Guide on Automated Decision-Making,[5] these guidelines also offer implementation guidance. These guidelines provide normative standards against which agencies, and in some cases the public, can evaluate their own practices.
2. **Prohibitions and moratoria:** Some jurisdictions have banned or prohibited the use of particular kinds of 'high risk' algorithmic systems. In some cases, such as in Morocco's facial recognition policy, prohibitions are framed as temporary moratoria, which are intended to lapse once appropriate safeguards and accountability mechanisms are designed and implemented.[6] Prohibitions and moratoria have been most prominently applied to facial recognition technologies used by law enforcement, and in some cases local governments in the USA, including Portland,[7] Oakland[8] and San Francisco.[9]

---

4 Department for Digital, Culture, Media and Sport (2021). *Data Ethics Framework*. UK Government. Available at: www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

5 Commonwealth Ombudsman (2019). *Automated Decision-Making Better Practice Guide*. Government of Australia. Available at: https://www.ombudsman.gov.au/publications/better-practice-guides/automated-decision-guide

6 National Control Commission for the Protection of Personal Data, Morocco. (2020) 'Press release of 30/03/2020: Press release accompanying the publication of deliberation No. D-97-2020 du 26/03/2020'. (In French) Available at: https://www.cndp.ma/fr/presse-et-media/communique-de-presse/661-communique-de-presse-du-30-03-2020.html

7 City of Portland, Oregon. (2020). *City Council approves ordinances banning use of face recognition technologies by City of Portland bureaus and by private entities in public spaces*, Available at: https://www.portland.gov/smart-city-pdx/news/2020/9/9/city-council-approves-ordinances-banning-use-face-recognition

8 City of Oakland, California. Chapter 9.64 – Regulations on city's acquisition and use of surveillance technology, Title 9 – Public Peace, Morals and Welfare, *Oakland, California Code of Ordinances*. Available at: https://library.municode.com/ca/oakland/codes/code_of_ordinances?nodeId=TIT9PUPEMOWE_CH9.64REACUSSUTE

9 City and County of San Francisco. (2019). Chapter 19B: Acquisition of Surveillance Technology, *San Francisco Administrative Code*. Available at: https://codelibrary.amlegal.com/codes/san_francisco/latest/sf_admin/0-0-0-47320

3. **Public transparency:** Transparency mechanisms provide information about algorithmic systems to the general public (e.g. affected persons, media or civil society) so that individuals or groups can learn that these systems are in use, and demand answers and justifications related to such use. Examples of these transparency efforts include:

    a. public registries of algorithmic systems Ontario,[10] Amsterdam,[11] Helsinki,[12] and in cities in France, including Antibes, Lyon and Nantes[13] which are aimed at civil society and citizens

    b. requirements for source-code transparency, which apply to computational algorithmic systems, and have been implemented under the Canadian Directive on Automated Decision Making (ADM)[14]

    c. explanations of algorithmic logics (purportedly allowing the public and policymakers to 'understand' how an algorithmic decision was reached). This is a legal requirement under French law in the Digital Republic Bill.[15]

4. **Impact assessments:** Impact assessments include a broad range of accountability mechanisms that have been implemented in scientific and policy domains as wide-ranging as environmental protections, human rights, data protection and privacy. The goal is to mitigate harmful impacts of a given initiative or deployment, recognising risks and addressing them before implementation.

    Algorithmic impact assessments (AIAs) are mechanisms intended for public agencies to better understand, categorise and respond to the potential harms or risks posed by the use of algorithmic systems, usually *prior* to their use. AIAs vary substantially, but were originally recommended as a way to allow affected stakeholders to define and construct a matrix of harms, benefits and risks in order to evaluate *ex ante* whether the use of an algorithmic system is suitable in a particular context.

    It is purported that AIAs provide impacted communities in particular with more involvement in the uses of algorithmic systems by public agencies, and influence over how they respond to potential harms.[16] In practice, however, most AIAs currently in use have not engaged these communities substantively, and have been applied primarily for internal self-assessment by public agencies so are often not publicly available. In some cases, for example, under the Canadian Directive on Automated Decision-Making,[17] or the New Zealand Algorithm Charter,[18] the outcomes of AIAs go on to determine the eventual level of regulatory scrutiny applied to particular algorithmic systems.

---

10 Ontario. *Data Catalogue*. Available at: https://data.ontario.ca/group/artificial-intelligence-and-algorithms

11 City of Amsterdam Algorithm Register Beta. *What is the Algorithm Register?* Available at: https://algoritmeregister.amsterdam.nl/en/ai-register/

12 City of Helsinki AI Register. *What is AI Register?* Available at: https://ai.hel.fi/en/ai-register/

13 Pénicaud, S. (2021). 'Building Public Algorithm Registers: Lessons Learned from the French Approach'. *Open Government Partnership Blog*. 12 May Available at: https://www.opengovpartnership.org/stories/building-public-algorithm-registers-lessons-learned-from-the-french-approach/.

14 Treasury Board of Canada Secretariat, Government of Canada. (2019). *Directive on Automated Decision-Making*. Available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

15 Republique Francaise. (2016). *The Digital Republic bill – Overview*. Available at: https://www.republique-numerique.fr/pages/in-english

16 Metcalf, Jacob, et al. (2021). 'Algorithmic impact assessments and accountability: The co-construction of impacts.' *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Available at: https://dl.acm.org/doi/abs/10.1145/3442188.3445935

17 Treasury Board of Canada Secretariat, Government of Canada. (2019). *Directive on Automated Decision-Making*. Available at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

18 New Zealand Government. (2020). *Algorithm Charter for Aotearoa New Zealand*. Available at: https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

5. **Audits and regulatory inspection:** Audits refer to a range of mechanisms which are intended to provide insight into the functioning and potential impacts of an algorithmic system. For this report, we use two relevant definitions of 'audit,' as outlined in Ada's Examining the Black Box report:

    a. **Technical audit:** A narrowly targeted test of a particular hypothesis about a system by looking at its inputs and outputs – for instance, seeing if it exhibits racial bias in the outcomes of a decision.

    b. **Regulatory audit:** In this context an audit refers to a regulatory inspection and compliance exercise, such as a financial audit. Increasingly, regulatory inspections are also designed to capture the broader social consequences of a system's use, and assess its functioning with respect to an established normative standard, in order to identify potential areas of concern.[19]

    Technical audits and regulatory inspections can vary in their scope and application, but in general, they rest on the assumption that inspections help create an independent account of how algorithmic systems function, and account for any flaws, biases or bugs in the system.

    While audits are an important mechanism for public sector accountability,[20] and in combination with other approaches hold promise for algorithmic systems, they have not been formalised as standard policy mechanisms for public sector use of algorithmic systems. To date, they remain largely ad-hoc exercises conducted under the wider ambit of particular regulatory or administrative agencies, including statutory auditors in Sweden,[21] the Netherlands[22] and in France.[23] The UK's Information Commissioner's Office has also encouraged internal regulatory auditing by organisations using artificial intelligence, including both compliance audits as well as technical 'bias' audits, in its draft Guidance on the AI Auditing Framework.[24]

19 Ada Lovelace Institute and DataKind UK. (2020). *Examining the Black Box: Tools for Assessing Algorithmic Systems*. Available at: https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems

20 Supreme Audit Institutions of Finland, Germany, the Netherlands, Norway and the UK. (2020). *Auditing machine learning algorithms: A white paper for public auditors*. Supreme Audit Institutions of Finland, Germany, the Netherlands, Norway and the UK. Available at: https://www.auditingalgorithms.net/

21 Swedish National Audit Office. (2020). *Automated Decision-Making in Public Administration*. Available at: https://www.riksrevisionen.se/en/audit-reports/audit-reports/2020/automated-decision-making-in-public-administration---effective-and-efficient-but-inadequate-control-and-follow-up.html

22 Netherlands Court of Audit. (2021). *Understanding algorithms*. Available at: https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms

23 Cour des Comptes (Court of Auditors). (2017). *Admission post-bac and access to higher education*. Available at: https://www.eurosai.org/fr/databases/audits/Admission-post-bac-and-access-to-higher-education/

24 UK Information Commissioner's Office. (2020). *Draft Guidance on the AI Auditing Framework*. Available at: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

6. **External/independent oversight bodies:** Independent oversight mechanisms are intended to ensure accountability by monitoring the actions of public bodies, and making recommendations, sanctions or decisions about their use of algorithmic systems. Oversight mechanisms vary widely in form and function. Some mechanisms rely upon legislative oversight, as in Community Control of Police Surveillance legislation in the USA.[25] Others, like the Algorithm Management and Policy Officer in New York City,[26] are implemented through executive offices. Others, such as the West Midlands Police Data Ethics Committee, function in advisory capacities without specifically delegated legal powers.[27]

7. **Rights to hearing and appeal:** Some policies require that decisions made with the aid of algorithmic systems adhere to particular procedures, as a means of ensuring fairness and providing forums for individual redress in the case of a biased or erroneous decision. These procedures, which include notice of the decision, the provision of a hearing, the ability to present evidence and/or the right to appeal a decision to a neutral forum, are intended to provide forums for affected individuals or groups to debate or contest particular decisions that affect them. The most prominent of these are requirements of notice, hearing and rights to explanation of automated decisions provided under the GDPR in the EU.[28]

8. **Procurement conditions:** Government procurement conditions have been an important area of intervention for transparency and accountability.[29] Some policies attempt to translate these general rules of transparency and accountability to algorithmic systems. When governments acquire algorithmic systems from private vendors, particular procurement conditions may be applied that limit the design and development of an algorithmic system (e.g. to ensure that a system considered for procurement is transparent and non-discriminatory).

   These contractual pre-conditions are meant to ensure that governments only acquire systems that comply with transparency, fairness or other requirements, and that, in case a vendor fails to meet conditions, the vendor is subject to contractual liability. Procurement conditions have been established as policy mechanisms by the City of Amsterdam in the Netherlands,[30] have been promoted by the UK Government through its guidelines for AI procurement,[31] and the state government of Tamil Nadu, in India, in its policy on the safe and ethical use of AI.[32]

---

25 ACLU. (2021). *Community Control Over Police Surveillance (CCOPS) model bill*. Available at: https://www.aclu.org/legal-document/community-control-over-police-surveillance-ccops-model-bill

26 Office of the Mayor, City of New York. (2019). *Executive order no. 50: establishing an algorithms management and policy officer*. Available at: https://www1.nyc.gov/assets/home/downloads/pdf/executive-orders/2019/eo-50.pdf

27 West Midlands Police and Crime Commissioner. (2021). *Ethics Committee*, Available at: https://www.westmidlands-pcc.gov.uk/ethics-committee/

28 Article 29 Working Party. (2018). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. https://ec.europa.eu/newsroom/article29/items/612053/en

29 Open Government Partnership. *Open Contracting and Public Procurement*. Available at: https://www.opengovpartnership.org/policy-area/open-contracting/

30 Municipality Amsterdam. (2020). *Standard Clauses for Municipalities for Fair Use of Algorithmic Systems*. Available at: https://www.amsterdam.nl/innovatie/

31 UK Government. (2020). *Guidelines for AI Procurement*. Available at: https://www.gov.uk/government/publications/guidelines-for-ai-procurement

32 Government of Tamil Nadu. (2020). *Safe and Ethical AI Policy*. Available at: https://elcot.in/sites/default/files/AIPolicy2020.pdf

Learning from the first wave of policy implementation

This report does not set out to definitively evaluate the effectiveness of particular algorithmic accountability policies in all settings. The policies we review here are relatively new (concentrated within the last two to three years), making it difficult to assess their intermediate or long-term effects. Abstract findings of effectiveness will have little value in situated local or national contexts, though we do encourage future studies of holistic effectiveness as these practices continue to mature and increase in adoption.

Based on our review of the first wave of algorithm accountability policies, this report identifies six factors, which act as key determinants for the effective deployment and implementation of algorithmic accountability policies.

1. Clear institutional incentives and binding legal frameworks can support consistent and effective enforcement of accountability mechanisms, supported by reputational pressure from media coverage and civil society activism

Institutional and legal structures are important factors shaping the possibilities and practical dimensions of the implementation of algorithmic accountability policies. Enabling legal frameworks can provide important incentives to operationalise algorithmic accountability policies within public agencies that use algorithmic systems.[33]

Such frameworks include sanctions for non-enforcement, as well as the institutionalisation of policy mechanisms by embedding them within existing structures of government accountability, for example, by requiring public notice and comment processes for policy-level decisions made using algorithmic systems, or extending administrative due process frameworks for decisions made about individuals with the use of algorithmic systems. These frameworks are also crucial in empowering external oversight bodies as well as audits or regulatory inspection.

That said, the likelihood of establishing a legal framework (which, typically, involves a protracted legislative process) is vulnerable to the vagaries of changing political will for enacting and enforcing legal commitments, and is not something eager public servants can simply will into being. Indeed, in this first wave of policy implementation, most government agencies have

33 Bovens, M., and A. Wille. (2020). 'Indexing Watchdog Accountability Powers, A Framework for Assessing the Accountability Capacity of Independent Oversight Institutions.' *Regulation and Governance*. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/rego.12316

experimented with policymaking outside of the legislative process, implementing these systems as voluntary measures or executive decrees without a legal backstop. The hope is that these voluntary frameworks can set the foundation for a broader push for legal commitments in the future. In New Zealand, for example, a prominent civil society demand is that the Algorithm Charter (the country's voluntary policy) is now enacted in legislation.

Aside from legal frameworks, internal institutional incentives within government agencies also play a role in implementation. Notably, cultural norms of public service and reputational risks from media coverage and advocacy campaigns have functioned as important incentives for the implementation of policy mechanisms governing the use of algorithmic systems. Public officials we interviewed also noted that while public agencies want to be seen to be modernising and innovating in the AI and technology sector, reputational concerns have spurred an institutional focus on 'responsible innovation' (a term that appears to be an increasingly popular frame among public officials) and has encouraged the publication of information about how agencies are implementing policy commitments.

2. Algorithmic accountability policies need to clearly define the objects of governance as well as establish shared terminologies across government departments

Definitional ambiguity has been a key obstacle to both the design and the implementation of algorithmic accountability policies within public agencies. For example, policies vary in choice and interpretation of terms – referring to the object of governance variously as 'algorithms,' 'automated decision systems,' and 'data science/analytics.'

The lack of standard practice or shared vocabularies about the underlying object of governance is responsible for substantial confusion within public agencies regarding their need to comply with policy mechanisms, and how to execute such compliance.[34] This ambiguity was also felt by agencies tasked with designing appropriate policy responses and those responsible for oversight, for example, by the members of the New York Automated Decision Making Task Force.[35]

Some policies, like the New Zealand Algorithm Charter,[36] choose a broad definition that does not stipulate rigid technical thresholds, instead recognising that even relatively simple algorithmic systems can cause failures of accountability and harm. Further, some policies adopt a definition focused on impact and use, encompassing both the technologies as well their contexts of use, from there focussing on the situated uses of a technology that raise cause of concern. For example, the GDPR and the Canadian Directive on Automated Decisions identify 'automated decision-making' as the underlying object of regulation, in recognition of the concerns caused by technologies that automate, aid or replace human decision-making.

---

34 Richardson, R., (2021). 'Defining and Demystifying Automated Decision Systems' *Maryland Law Review*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3811708

35 See Richardson, above.

36 New Zealand Government. (2020). *Algorithm Charter for Aotearoa New Zealand*. Available at: https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

Regardless of technological form, each policy's purported intent is to ensure that algorithmic systems can be held to account in ways similar to human decision-making. As such, definitions should focus on technological systems or processes as a whole, applied in contexts where they 'automate, aid, or replace' human decision-making. Adopting broad definitions, particularly in an area where new accountability concerns are constantly being unearthed, can also ensure much-needed dynamism in the application of policy mechanisms. Rashida Richardson has helpfully provided the following model definition of an 'automated decision system', which might be suitably adapted to sector-specific concerns:

> *'"Automated Decision Systems" are any systems, software, or process that use computation to aid or replace government decisions, judgments, and/or policy implementation that impact opportunities, access, liberties, rights, and/or safety. Automated Decisions Systems can involve predicting, classifying, optimizing, identifying, and/or recommending.'*[37]

3. Setting the appropriate scope of policy application supports their adoption. Existing approaches for determining scope such as risk-based tiering will need to evolve to prevent under- and over-inclusive application

Given limited public resources, defining the scope of policies – how they're applied and to what – allows policymakers to identify priority areas of concern or impact. A number of factors contribute to how policymakers define the scope of policy application – including the perceived risk presented by the use of a particular algorithmic system in certain contexts (for example, those used within law enforcement).

Despite its intuitive appeal in narrowing the scope of policy application, binary or rigid risk-tiering (such as 'high-risk' versus 'no-risk') might also result in certain technologies being wrongfully excluded from or included in certain tiers. Assessments of risk should be appropriate to the specific social, political and institutional contexts within which an algorithmic system is being deployed. For example, certain 'high-risk' contexts might already apply institutionalised forms of accountability which are responsive to their particular contexts, even if these policies don't mention 'algorithms' specifically. In other instances, what might otherwise be perceived as a 'low-risk' use-case for algorithmic systems might exist in a context of less scrutiny or institutional accountability.

In all cases, risks and harms are ultimately contextual and, in many circumstances, defy measurement – assessing 'high' vs 'low-risks' and harms will depend on which perspectives are included in an assessment, who is doing the assessing, and which framework is used. A potential benefit for some parties will be a risk or harm to others.

---

37 Richardson, R., (2021). 'Defining and Demystifying Automated Decision Systems' *Maryland Law Review*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3811708

Addressing this challenge requires a regulatory design that balances a risk-based approach with methods for resolving uncertainty that involve monitoring and addressing novel threats caused by algorithmic systems. Such an approach also demands greater precaution and recognition of systematic and diffuse harms that might arise through the use of algorithmic systems, but may be difficult to quantify and manage in risk-based approaches.[38]

In some cases policies are limited in their application to algorithmic systems which concern individuals only, such as the GDPR and the Canadian ADM Directive. In the case of the GDPR and some of its national implementations,[39] the application of some rules is also limited to algorithmic decisions which are 'solely automated' and do not involve human intervention. This is a limitation which introduces substantial ambiguity in application[40] and ignores complex interactions of algorithmic systems and human decision-makers.[41]

Other policies have a broader focus and recognise not only individual impact, but also the impact of algorithmic systems on particular groups. These include the New Zealand Algorithm Charter and the UK Data Ethics Framework. While those we spoke to in government were also cognisant of the broader societal impacts that the use of algorithmic systems might have on the transparency and accountability of policymaking or administrative processes, the policies reviewed in this report did not explicitly include this concern within their scope, which is a noticeable gap.

4. Policy mechanisms that focus on transparency must be detailed and audience appropriate to underpin accountability

Meaningful transparency is an expectation and explicit mechanism in a number of policy interventions for algorithmic accountability. In particular, accountability is dependent on the public's ability to know that algorithmic systems are being used in the first place, to assess information about the systems and to demand responses about their use.[42] Respondents identified two broad transparency concerns related to the operation of algorithmic systems, and to the implementation of accountability policies themselves.

---

38 Cohen, Julie E. (2016). 'The regulatory state in the information age.' *Theoretical Inquiries in Law* 17.2: 369-414. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2714072

39 Malgieri, G., (2019). 'Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations.' *Computer Law & Security Review* Vol. 35 No.5. Available at: https://www.sciencedirect.com/science/article/pii/S0267364918303753

40 Privacy International. (2017). *Data Is Power: Profiling and Automated Decision-Making in GDPR*. Available at: https://privacyinternational.org/sites/default/files/2018-04/Data%20Is%20Power-Profiling%20and%20Automated%20Decision-Making%20in%20GDPR.pdf

41 Green, B and Chen, Y. (2019). 'Disparate Interactions: An Algorithm-inthe-Loop Analysis of Fairness in Risk Assessments'. In FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19), January 29–31, 2019, Atlanta, GA, USA. ACM, New York, NY, USA. Available at: https://doi.org/10.1145/3287560.3287563

42 Kemper, J., and Kolkman, D., (2019). 'Transparent to whom? No algorithmic accountability without a critical audience.' *Information, Communication & Society* Vol. 22. No. 14 pp. 2081-2096. Available at: https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1477967

First, transparency is in tension with countervailing policy objectives that require confidentiality, including issues of security, privacy, intellectual property concerns (particularly in the case of systems acquired from private vendors) and the risk of algorithmic systems being 'gamed' by adversaries.[43] For example, agencies implementing a fraud investigation algorithm may be wary of releasing the details of the algorithmic system and how it 'recognises' fraud, in the event that fraudulent actors may use this knowledge to try to circumvent the system. While there are convincing arguments for secrecy in some cases, there needs to be particular vigilance against broad and unsubstantiated claims that security risks will result from public disclosures. During deliberations for the (now enacted) New York City POST Act, experts deposed before the City Council testified that while the public disclosures about the City's surveillance systems demanded by the Act would provide valuable insights to the public, they were in no way detailed enough to enable someone to game the system and threaten public safety.

Second, different communities have different needs, and demand different mechanisms for transparency and accountability. This requires government agencies to be more cognisant of what kinds of information is being made available, and how particular audiences may make use of, rely upon, or gain access to it in the first place.[44] For example, the kinds of accountability enabled by releasing source code would be different from what would be afforded by providing plain language explanations of how a system works to impacted communities. Some mechanisms for transparency, like algorithm registers in Amsterdam and Helsinki, were specifically designed with critical audiences from civil society in mind, while in France, plain language audio-visual explanations of algorithmic systems were seen as important means of reaching impacted communities and the public.

Finally, the lack of transparency about the functioning of algorithmic accountability policies themselves can be a barrier to ensuring effective accountability. Most policies are not designed in a way that considers and accommodates the need to publicly communicate the outcomes of a given policy mechanism, or to update the public on processes like impact assessments and audits.[45] However, public communication of outcomes can be an important factor in enabling stakeholders to scrutinise the effectiveness of algorithmic accountability policy mechanisms. And some policies, like the Canadian Directive on ADM, are pointing in this direction, stipulating public transparency around the use of particular mechanisms like Algorithmic Impact Assessments.

43 Freeman Engstrom, D., et. al. (2020). *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Available at: https://law.stanford.edu/education/only-at-sls/law-policy-lab/practicums-2018-2019/administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/acus-report-for-administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/

44 Ananny, M., and Crawford, K., (2016). 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.' *New Media & Society* vol. 20 no.3 pp. 973-989. Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444816676645?journalCode=nmsa

45 Kaminski, M. E., and Malgieri, G., (2020). 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations.' *International Data Privacy Law* Available at: https://dl.acm.org/doi/abs/10.1145/3351095.3372875

5. Public participation supports policies that meet the needs of affected communities. Policies should prioritise public participation as a core policy goal, supported by appropriate resources and formal public engagement strategies

Civic participation and public engagement refers to the ability of diverse groups of stakeholders – including affected persons, community organisers, civil society organisations, public officials, or the general public – to participate in the design and implementation of algorithmic accountability policies. These groups provide diverse forms of grounded expertise and experience which policymakers and algorithmic system vendors may not otherwise have access to. This is especially crucial given that many automated systems being used in the public sector interface directly with individuals and groups, and impact their lives in tangible ways. Participation can shape the substance and forms that algorithmic accountability mechanisms take, as well as impacting their application, ensuring that citizens are more actively engaged with policy implementation.[46]

Most algorithmic accountability policy mechanisms we surveyed did not have clear and formal mechanisms for public engagement during their design or implementation. Forums of public engagement, where they did exist, were mostly limited to consultations with specific groups of stakeholders, including public officials responsible for the implementation of policy commitments in specific agencies. That said, some jurisdictions did incorporate participation either as a foundational principle (as in the case of the New Zealand Algorithm Charter), or through mandated consultation procedures (as in the case of some local surveillance control laws as in Oakland or Seattle).

Effective public engagement requires that policymakers first identify the kinds of diverse expertise needed to design and implement policy mechanisms, and enable this expertise to be communicated to policymakers through formal channels of engagement.[47] These include consultations, formal hearings, or more direct forms of democratic control through systems of legislature. Meaningful participation also requires ensuring that stakeholders are able to participate by ensuring that adequate time and resources are allocated in the process of engagement.[48]

---

46 Pallett, H., Burall, S., Chilvers, J. and Price, C. (2021). 'Public Engagement with Algorithms in Public Services' 3S Research Group Briefing Note. Available at: https://uea3s.files.wordpress.com/2019/10/public-engagement-with-algorithms-in-public-services.pdf

47 Richardson, R (ed.). (2019). '*Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*'. AI Now Institute. Available at: https://ainowinstitute.org/ads-shadowreport-2019.pdf

48 Ibid.

6. Policies benefit from institutional coordination across sectors and levels of governance to create consistency in application and leverage diverse expertise

Algorithmic systems are often complex and operate in dynamic environments. Various government or non-government agencies might be involved in the design and deployment of a single algorithmic system, which can make the attribution of responsibility and effective coordination between agencies challenging. This challenge is magnified in the case of an algorithmic system licensed from a vendor. Such confusion could give rise to fragmentation in the application of accountability policies, for example, where different components of an algorithmic system are subcontracted for procurement, and potentially fall out of the scope of transparency and accountability conditions. As such, governments should carefully consider the range of actors who should be the focus of monitoring and compliance for algorithmic accountability policies.

Moreover, the implementation of algorithmic accountability policies often requires governments to leverage diverse forms of expertise, including experience in IT systems, law, public administration, data science and statistics, data protection and privacy. The Canadian Government's AIA framework, for instance, requires the agencies implementing this policy to answer questions relating to legal frameworks, IT systems, data science and data protection. Coordination between various agencies or public officials with different kinds of expertise needs to be factored into policy interventions in these complex environments.[49]

Even though policy implementation can require diverse forms of expertise, particular agencies are often tasked with overseeing and coordinating the design and implementation of algorithmic accountability policies across government agencies. For example, Etalab is an open data task force responsible for the implementation of algorithmic accountability policy in France, while in Canada, the Treasury Board Secretariat has specific functions related to implementing the ADM Directive. It is necessary to examine the functioning of these coordinating institutions in order to understand how different institutional priorities and agendas may affect the implementation of policy goals, as well as how they are expected to coordinate between multiple agencies and stakeholders.

Algorithmic accountability policy is also affected by conflicts between policymaking at a global level, and its implementation at the local level. One way in which this challenge affects implementation is in how knowledge about policy mechanisms and requirements framed at a global level reaches the intended audiences within public agencies responsible for their implementation. Another challenge arises in ensuring that norms of accountability framed at a global level can be flexible enough to take into account the diverse and context-specific considerations that might be required for ensuring effective implementation at the local level, for example, in how parameters of fairness and discrimination might change according to context.

49 Treasury Board of Canada Secretariat, Government of Canada. (2019). *Algorithmic Impact Assessment Tool*. Available at: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

Conclusions and priorities for future research

Despite the limitations of research in an emergent area of policy, our research nonetheless indicates some of the challenges and key areas that should guide the design and implementation of algorithmic accountability policies as they develop, as well as pointing to broader considerations for the policymaking process:

For policymakers

As our mapping of algorithmic accountability policy indicates, most of these policies have emerged very recently (largely from 2019 onwards), and discussions about algorithmic accountability policy, as well as their implementation, have been concentrated in a few jurisdictions. This research, therefore, draws from this current narrow pool of evidence – acknowledging that the landscape of algorithmic accountability could look very different as recognition of the challenges posed by algorithmic systems in the public sector grows.

Our review of publicly available literature, as well as insights gathered from qualitative interviews and workshops with policy implementers and civil society, indicates that evidence about the impact and effectiveness of algorithmic accountability policies is currently limited. This is not surprising for a nascent field, but creates an opportunity for policymakers to integrate practices that enable policy monitoring and evaluation early on. These practices should include:

- Systematic and effective transparency around assumptions and objectives behind policy mechanisms.
- Strategic engagement with sectoral agencies.
- Timely and ongoing engagement with affected communities and civil society.
- Leverage peer-to-peer learning networks and global principles.

For the research community

This study also draws from, and builds on diverse scholarship that studies algorithmic systems with a view to promoting fairness, accountability and transparency in their design and use, particularly in the public sector. Our review of literature in this field indicates that the evidence base for algorithmic accountability research in the public sector could be strengthened, and that future research should prioritise the following areas of study:

- Further study and trials of algorithmic accountability approaches in practice.
- Continuous evaluation of algorithmic accountability policies.
- Widening research contexts to be more inclusive of diverse regional and institutional experiences.

Project team

Project Leads: Tonu Basu is the Deputy Director of Thematic Policy Areas at the Open Government Partnership, Jenny Brennan is a Senior Researcher at the Ada Lovelace Institute and Amba Kak is the Director of Global Policy & Programs at the AI Now Institute at New York University.

Lead Researcher: Divij Joshi is a lawyer and researcher interested in the social, political and regulatory implications of emerging technologies and their intersections with human values.

About the partners

For the **Ada Lovelace Institute (Ada)**, this research forms part of their wider work on algorithm accountability and the public sector use of algorithms. It builds on existing work on tools for assessing algorithmic systems, mechanisms for meaningful transparency on use of algorithms in the public sector, and active research with UK local authorities and government bodies seeking to implement algorithmic tools, auditing methods and transparency mechanisms.

For the **AI Now Institute**, law and policy mechanisms are a key pathway toward ensuring that algorithmic systems are accountable to the communities and contexts they are meant to serve. This research builds upon a wider body of work including their framework for Algorithmic Impact Assessments (AIA) and the Algorithmic Accountability Toolkit. In the spirit of proactive engagement with the policy process, alongside a broad civil society coalition, they also published the Shadow Report to the New York City Automated Decision Systems (ADS) Task Force to detail accountability mechanisms for various sectors of the city government.

For the **Open Government Partnership (OGP)**, a multi stakeholder partnership of 78 countries and 76 local jurisdictions, transparency, accountability and participation are key approaches to better policy making. OGP members work with civil society and other key actors in their countries to co-create and implement OGP action plans with concrete policy commitments, which are then independently monitored for ambition and completion through the OGP's Independent Reporting Mechanism. While several OGP countries are implementing their digital transformation agenda through their engagement in OGP, a growing number of OGP members are also using their OGP action plans to implement policies that govern public sector use of digital technologies. Among these, accountability of automated decision-making systems and algorithms has seen increasing interest. OGP convenes an informal network on Open Algorithms with implementing governments, mobilising a cross-country coalition of those working on algorithmic accountability.

Acknowledgements

This project was made possible due to the efforts and contributions of a number of people. We would like to thank the interviewees, as well as the participants at the workshops and discussions organised by the Open Government Partnership for participating in this study and lending their time and experience to this report—in particular the Governments of Canada, France, New Zealand, Netherlands, and the UK.

We would also like to thank Paula Pérez, José Perez Escotto, Joseph Foti, Helen Turek, Alan Wu, Imogen Parker, Andrew Strait, Octavia Reeve, Hannah Kitcher, Meredith Whittaker, Alejandro Calcaño, and Luke Strathmann for their feedback and assistance on this project.