Ada Lovelace Institute

AI NOW INSTITUTE

Open Government Partnership

**Pre-release executive summary:**
Algorithmic accountability for the public sector: learning from the first wave of policy implementation

**Date:**
Friday 4 June 2021

# Algorithmic accountability for the public sector: learning from the first wave of policy implementation

## Executive summary

The Ada Lovelace Institute (Ada), the AI Now Institute (AI Now), and the Open Government Partnerships (OGP) are partnering to launch the first global study to analyse the initial wave of algorithmic accountability policy. This pre-release Executive Summary was put together for the June 8 RightsCon Session titled "Inside the Black Box: Making Public-Sector Algorithms Accountable to Citizens and Communities." This project aims to understand the challenges and successes of algorithmic accountability policies by focusing on the experiences of the actors and institutions directly responsible for their implementation on the ground.

## Introduction and scope of study

Governments are increasingly turning to algorithms to automate or support decision-making for public services. Algorithms might be used to predict future criminals, make decisions about welfare entitlements, detect unemployment fraud, decide where to send police, or assist in urban planning. Yet growing evidence suggests that these systems can cause harm and frequently lack transparency in their implementation, including opacity around decisions about whether and why to use them. Furthermore, 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.

With many challenges and open questions arising from their early stages of implementation, Ada, AI Now, and OGP have partnered to launch the first global study to evaluate this initial wave of algorithmic accountability policy. This report presents and analyses evidence on the implementation of algorithmic accountability policies in different contexts, while also exploring the limits of legal and policy mechanisms in delivering meaningful redress. Our aim is to provide practical guidance to the policymakers, civil society, public officials, and agencies responsible for implementing related policy tools and commitments, as well as outline future directions for the lively research community in this field.

## Methodology

This report is based on a database we built of more than 40 examples of algorithmic accountability policies at various stages of implementation, from more than 20 national and local governments. We also conducted 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 City and New York City. We presented preliminary findings and received feedback at a workshop with countries that are part of the Informal Network on Open Algorithms convened by Open Government Partnership (OGP), with countries implementing commitments focusing on algorithmic accountability through their OGP action plans. We also reviewed 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, evaluating or analysing these policies without recognizing the contexts in which they are implemented tends to be inadequate.

The focus of this study is on understanding and analysing the factors that have shaped the implementation of algorithmic accountability policies in various jurisdictions. Consequently, instead of offering prescriptions, or rigid normative evaluations of particular policies in the abstract, we sought to describe and analyse how these factors may operate in particular contexts, and how they might enable or disable the objectives that these policies set out to achieve.

## Key definitions

- **Algorithms and Algorithmic Systems:** For the purpose of this report, we use the term 'algorithms' to describe a set of technologies used to computationally generate knowledge or decisions, operate on particular datasets, and that are bounded by specific logics and procedures.

  An algorithmic system is a system that uses automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans. The term 'algorithmic system' often refers to a particular piece of software: for example, a computer program that takes as its input the school choice preferences of students and outputs school placements. Algorithmic systems are often conceptualised as neutral or machine-driven, so it's important to remember that they are designed by humans and involve some degree of human decision-making in their operation. Humans are also ultimately responsible for how a system receives its inputs (e.g. who collects the data that feeds into a system), how the system is used, and how a system's outputs are interpreted and acted on. In our analysis, we consider the technical as well as social, cultural, legal and institutional contexts where algorithms are embedded, and identify the important determinants for understanding the accountability of these systems to people they impact.

- **Accountability:** There is an established body of policy evidence and scholarship on open government and accountability in the public sector, which typically includes policy interventions such as standard setting, answerability, investigation and sanction or enforcement. For the purpose of this report, we use the term accountability to convey the set of efforts oriented towards ensuring that those that build, procure, and use algorithms are eventually answerable for their impacts. Beyond law and policy, accountability may include tech-worker organising and whistleblowing, community organisers, civil society organisation, and investigative journalism, all of which have emerged to hold these systems accountable to the contexts and communities they are meant to serve.

  Algorithmic accountability is also a growing interdisciplinary field of research that focuses on fairness, accountability, and transparency in algorithmic systems, with contributions from computer science, the social sciences, and law.

## A typology of algorithmic accountability policies

To understand the impact, implementation and effectiveness of algorithmic accountability policies, we first look at their intentions, aims, and assumptions. 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. As a relatively recent field, these policies vary widely.

A shared vocabulary of 'algorithmic accountability policy' is still being built, and as such, definitional boundaries are blurry and shifting. The following typology, while not intended to be comprehensive, indicates the forms of algorithmic accountability policies taking shape in the public sector.

1. **Principles and guidelines:** A number of policy documents provide normative guidance on 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 in the use of algorithmic

systems by public agencies. In some cases, as in the UK Data Ethics Framework, or the Australian Better Practice Guide for Automated Decision-Making, these guidelines also offer implementation guidance. While not binding, these guidelines provide normative standards against which agencies can evaluate their own practices.

2. **Prohibitions and moratoriums:** Some jurisdictions have banned or prohibited the use of particular kinds of algorithmic systems which are perceived as being 'high risk'. In some cases, such as in Morocco's facial recognition policy, prohibitions are framed as temporary moratoriums, which are intended to lapse when appropriate safeguards and accountability mechanisms can be designed and implemented. Prohibitions and moratoriums have been most prominently applied to facial recognition technologies used by law enforcement, or in local governments in the USA, including Oakland and San Francisco.

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 demand answers and justifications for the use of those systems. Examples of these transparency efforts includes: Public registries of algorithms in Amsterdam, Helsinki, Nantes, Antibes and Ontario, which are aimed at civil society and citizens; Requirements for source code transparency, which have been implemented under the Canadian Directive on ADM; and Explanations of algorithmic logics, which is a legal requirement under French law in the Digital Republic Bill.

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. 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 generally aim 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. AIAs, in general, are intended to provide

particularly impacted communities, more insight into the uses of algorithmic systems by public agencies, and how to respond to potential harms. In practice, however, AIAs currently in use have mostly been relied on internally for self-assessment by public agencies. In some cases, for example, under the Canadian Directive on Automated Decision-Making, or the New Zealand Algorithm Charter, the outcomes of AIAs determine the eventual level of regulatory scrutiny applied to particular algorithmic systems.

5. **Audits and regulatory inspection:** audits refer to a range of mechanisms which are intended to provide insight into the functioning of an algorithmic system. For this report, we use two relevant meaning of 'audit,' as outlined in Ada's Examining the Black Box report:
    a) **Audit from the perspective of the computer science community:** a narrowly targeted test of a particular hypothesis about a system by looking at its inputs and outputs – for instance, seeing if it has racial bias in the outcomes of a decision. In this report, this is called a bias audit.
    b) **Audit from the perspective of its use in regulatory contexts:** This is 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.

Audits and regulatory inspections can vary in their scope and application, but in general, rest on the assumption that inspections create an independent and objective account of how algorithmic systems function. While audits are an important mechanism for public sector accountability, particularly for machine learning systems, they have not been formalised as standard policy mechanisms for public sector algorithm use. They remain largely ad-hoc exercises conducted under the wider ambit of particular regulatory or administrative agencies, including statutory auditors in Sweden, the Netherlands and in France. 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.

6. **External 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. Others, like the Algorithm Management and Policy Officer in New York City, are implemented through executive offices. Others, such as the West Midlands Police Data Ethics Committee, function in advisory capacities without specifically delegated legal powers.

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. These procedures, which include notice of the decision, the provision of a hearing, the ability to present evidence, or the right to appeal a decision to a neutral forum, are intended to provide forums to affected individuals or groups to debate or contest particular decisions made about 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.

8. **Procurement conditions:** Government procurement conditions have been an important area of intervention for transparency and accountability. 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 implemented within the design and development of an algorithmic system (e.g., geared toward transparency or non-discrimination). These contractual pre-conditions ensure that governments only acquire systems that comply with transparency or fairness requirements, and that, in case a vendor fails to meet conditions, they are subject to contractual liability. Procurement conditions have been established as policy mechanisms by the City of Amsterdam in the Netherlands, and have also been promoted by the Government of UK through its Guidelines for AI Procurement, and the

state government of Tamil Nadu, in India, in its Policy on the Safe and Ethical use of AI.

## Learning from the first wave of policy implementation

For a number of reasons, this report does not set out to definitively evaluate particular algorithmic accountability policies - For example, abstract findings of effectiveness will have little value in situated local or national contexts. These policies are also relatively nascent (concentrated within the last 2-3 years), making it difficult to assess their intermediate or long-term effects. What these findings do illuminate, however, is that political, institutional, and contextual factors influence the effectiveness of certain algorithmic accountability policies. The report identifies six such factors as key determinants of effective policy implementation.

1. **Institutional incentives and legal frameworks**

Institutional and legal structures are important factors affecting the implementation of algorithmic accountability.

The legal framework within which policy mechanisms operate is important, although it is not a necessary nor sufficient determinant of effectiveness. Enabling legal frameworks can provide important incentives to operationalise algorithmic accountability policies within public agencies using algorithmic systems. These include sanctions for non-enforcement, as well as the institutionalization of policy mechanisms by embedding them within existing structures of government accountability. They are also crucial in empowering external oversight as well as audits or regulatory inspection.

That said, the likelihood of establishing a legal framework (which, typically, involves a protracted parliamentary process) is vulnerable to the vagaries of changing political will for enacting and enforcing legal commitments. In fact, in this first wave of policy implementation, most government agencies have

tried to experiment with policymaking outside of the legislative process, implementing these systems as voluntary measures or executive decrees.

Apart from legal frameworks, internal institutional incentives also play a role in effective implementation. Notably, cultural norms of public service within public agencies, reputational factors around the use of ethical innovation, and the risk of reputational injury caused by failure to comply with policy guidance, can provide important incentives for implementation of policy mechanisms.

2. **Defining the object of policy**

Definitional ambiguity has been a key obstacle to both the design and the implementation of algorithmic accountability policies. For example, policies vary in choice and interpretation of terms - referring to the object of governance 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. 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.

Some policies, like the New Zealand Algorithm Charter, choose a broad definition that does not stipulate rigid technical thresholds, instead recognising that even relatively simple algorithmic systems can cause failures of accountability. Further, some policies adopt a definition focused on impact and use, encompassing both the technologies as well their contexts of use, and focussing on the uses of a technology that are the 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.

Regardless of technological form, each policy's intent is to ensure that algorithms can be held to account similar to human decision-making. As such, definitions should focus on technological systems or processes,

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.

### 3. Defining the scope of application

Given limited public resources, defining the scope of policies allows policymakers to identify priority areas of concern or impact. A number of factors contribute to how policymakers define the scope of application - including the perceived risk presented by the use of a particular algorithmic system in certain contexts (for example, those used within law enforcement), or in recognition of the needs of certain stakeholders.
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, the application of some rules is also limited to decisions which are 'solely automated' and do not involve human intervention, a limitation which introduces substantial ambiguity in application and ignores complex interactions of algorithmic systems and human decision-makers.
Other policies have a broader focus and recognise not only individual impact, but also impacts on particular groups, including the New Zealand Algorithm Charter and the UK Data Ethics Framework. While governments were also cognizant 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 within their scope, which is a noticeable gap in the implementation of algorithmic accountability policy so far.

### 4. Enabling meaningful transparency

Meaningful transparency is an expectation and explicit mechanism of a number of policy interventions for algorithmic accountability. In particular, accountability is often tied to the public's ability to assess information and demand responses about the use of algorithmic systems. However, we've identified two broad transparency concerns related to the operation of algorithmic systems, as well as the implementation of accountability policies themselves.

First, transparency is affected by countervailing policy objectives requiring confidentiality, including concerns 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. For example, agencies implementing a fraud investigation algorithm may be wary of releasing the logic of the algorithm in the event that fraudulent actors may try to circumvent the system by changing their tactics.

Second, different communities have different needs, and demand different mechanisms for accountability. This requires government agencies to be more cognizant of what kind of information is being made available, and how particular audiences may make use of, or rely upon it. For example, the kinds of accountability enabled by releasing source code would be different from 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 keeping in mind critical audiences from civil society, 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 with consideration of the need to publicly communicate the outcomes of policy mechanisms or processes like impact assessments and audits - public communication of outcomes can be an important factor in enabling stakeholders to scrutinise the effectiveness of the mechanisms. However, some policies, like the Canadian Directive on ADM, do stipulate public transparency in the use of particular mechanisms like Algorithmic Impact Assessments.

### 5. Civic participation

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 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 accountability mechanisms take, as well as impacting their implementation, ensuring that citizens are more actively engaged with policy implementation.

Most of the policy mechanisms we surveyed for this report did not have clear and formal mechanisms for public engagement at the stage of their design or their implementation. Forums of public engagement 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 identify the kinds of diverse expertise needed to design and implement policy mechanisms, and enable this expertise to be filtered up to policymakers through formal channels of engagement. These include consultations, formal hearings, or more direct forms of democratic control through systems of legislature. Meaningful participation also requires ensuring that stakeholders are empowered to participate by ensuring that adequate time and resources are allocated in the process of engagement.

### 6. Institutional coordination

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 might 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 for procurement. 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. For example, the Government of Canada's AIA framework requires implementing agencies 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.

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, the Etalab in France 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. The functioning of these coordinating institutions should be carefully examined 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 a more decentralised, local level. One way in which this challenge affects implementation is in the ways that 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 contexts.

## Upcoming full report publication

This Executive Summary is part of a larger research effort and report that will be finalized and released publicly in Summer/Fall 2021.

The report will present an indicative overview of what could be called the 'first wave' of algorithmic accountability policies from around the world - some early responses from governments towards a challenge which is both dynamic and unstable. The report is intended to be a glimpse into the forms that government policy responses into algorithmic systems are taking, as well as into the factors that are shaping the implementation of these policies.

## About the partners

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 our framework for Algorithmic Impact Assessments (AIA), Algorithmic Accountability Toolkit and the Regulating Biometrics compendium. In the spirit of proactive engagement with the policy process, alongside a broad civil society coalition, we 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 Ada Lovelace Institute, algorithm accountability is one of four pillars of our work as a London-based research institute and deliberative body working to make data and AI work for people and society. This research forms part of our wider work on algorithm accountability. 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 using machine learning.

For the Open Government Partnership, a partnership of 78 countries and 76 local jurisdictions, advancing transparency and accountability in digital policy tools should be a critical part of any government agenda. 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. Several OGP members are implementing their digital transformation agenda through their engagement in OGP. A growing number of OGP members are 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 of public officials implementing algorithmic accountability reforms, including those initiated through their OGP action plans, serving as a cross-country coalition of those working on algorithmic accountability. This research seeks to serve as a resource for OGP members implementing commitments on this set of issues.

## 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, 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.