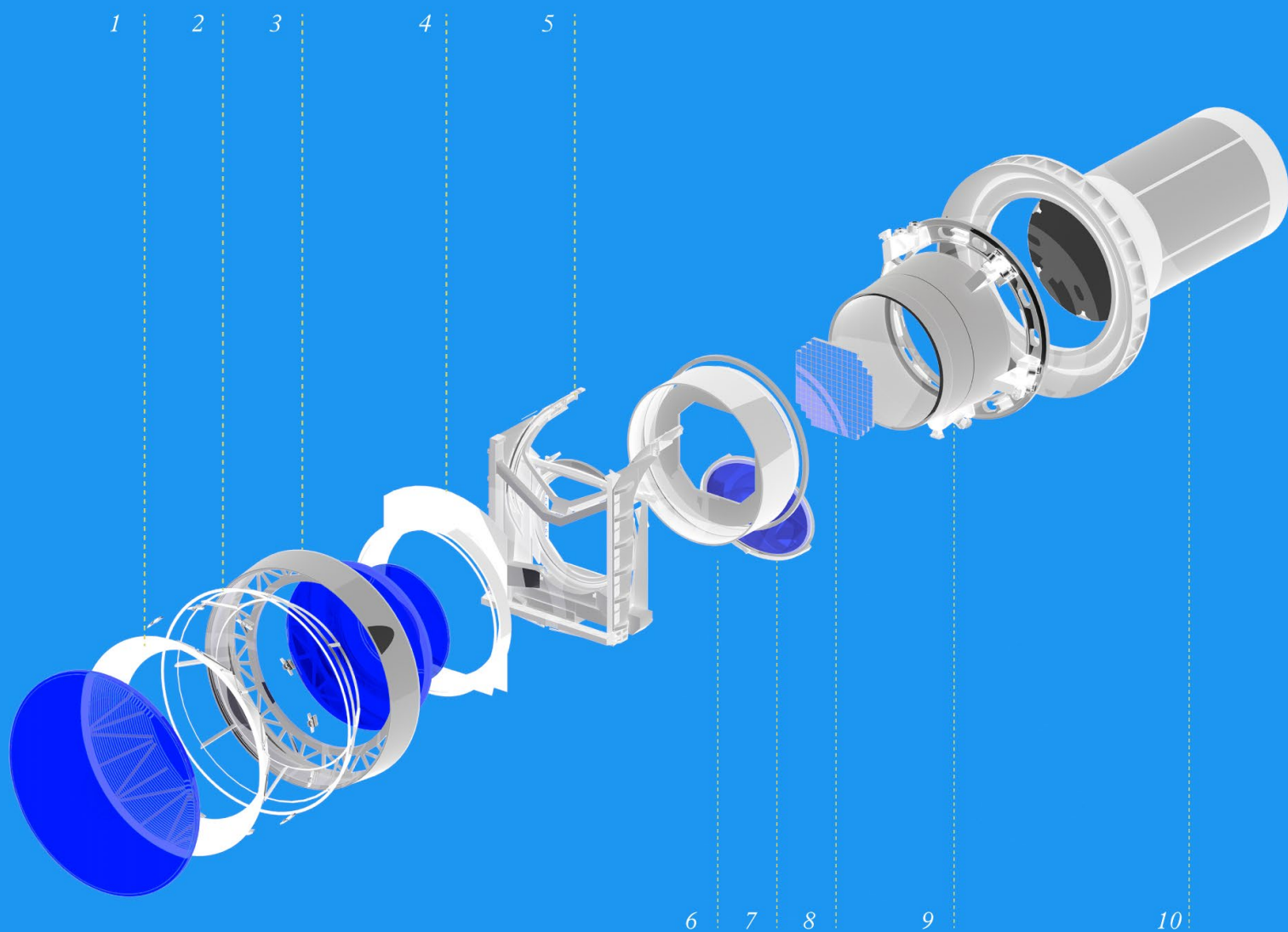


Assembling Accountability

Policy Brief

**DATA &
SOCIETY**

Emanuel Moss
Elizabeth Anne Watkins
Ranjit Singh
Madeleine Clare Elish
Jacob Metcalf



INTRODUCTION

We are at a turning point for the future of algorithmic accountability. Already, numerous jurisdictions have proposed legislation that would implement **Algorithmic Impact Assessment (AIA)** as a tool for bringing accountability to the algorithmic systems that increasingly permeate everyday life. Despite this heightened focus on impact assessments as an algorithmic governance mechanism, no one has yet created a fully realized and truly accountable AIA process. Our report, [Assembling Accountability](#), details what components remain unspecified by these efforts, and why it is necessary to foster a community of experts and advocates committed to building assessment practices in the public interest.

AIAs, or functionally similar processes, are core components of existing and proposed algorithmic regulations in the US, European Union, and Canada. But the final forms of these AIAs have yet to be defined; if organizations that build and deploy algorithmic systems are allowed to define for themselves how they are evaluated, many harms to the public interest will proceed unchecked.

In addition, algorithmic systems present a special challenge to assessors. The impacts of algorithmic systems are too often discovered only long after they have been integrated into society, and even then only through the work of a dedicated, but patchwork, network of watchdog technologists, auditors, journalists, and critical scholars. The harm these systems cause is often unevenly distributed, or is only visible in the aggregate. Because assessing algorithmic systems is inherently challenging, industry capture of the assessment practices is both dangerous and a likely outcome without concerted effort to incorporate the public interest.

To support policy conversations and ensure that the development of AIAs is effective, our report maps the challenges of constructing AIAs by analyzing impact assessments in other domains, including finance, environment, human rights and privacy. Building on this comparative analysis, **the report identifies ten constitutive components as a framework for evaluating current and proposed AIA regimes.**

We anticipate this framework can work as a practical tool for regulators, advocates, public-interest technologists, technology companies, and critical scholars who are identifying, assessing, and acting upon algorithmic impacts.

ALGORITHMIC IMPACT ASSESSMENT FRAMEWORK

Our research indicates that a singular, generalized model for AIAs would not be effective. Instead, this report illustrates the critical decision points in the development of AIAs and specifies ten necessary components that constitute robust impact assessment regimes. We propose that this new framework is needed to support the network of technologists, auditors, journalists, and critical scholars working in this space.

1. **Sources of Legitimacy:** Impact assessments need to be legitimized either through legislation or within a set of norms that are officially recognized and publicly valued.
2. **Actors and Forum:** Impact assessments are rooted in establishing an accountability relationship between actors that

design, deploy, and operate a system and a forum that can allocate responsibility for potential consequences of such systems and demand changes in their design, deployment, and operation.

3. **Catalyzing Event:** Such events are triggers for conducting impact assessments. These can be mandated by law or solicited voluntarily at any stage of a system's development life cycle.
4. **Time Frame:** Once triggered, the time frame is the period (often mandated through law or mutual agreement between actors and the forum) within which an impact assessment must be conducted. Most impact assessments are performed *ex ante*, before developing a system, but they can also be done *ex post* as an investigation of what went wrong.
5. **Public Access:** Achieving genuine transparency and accountability requires the ability of the public to scrutinize and contest an impact assessment's process and documentation. The broader the public access, the stronger is its potential to enact accountability.
6. **Public Consultation:** The conditions for solicitation of feedback should be from the broadest possible set of stakeholders in a system. Who constitutes this public and how they are consulted are critical questions for the success of an impact assessment.
7. **Methods:** These are standardized techniques of evaluating and foreseeing how a system would operate in the real world.

Most impact assessments have a roster of well-developed quantitative and qualitative techniques that can be applied to foresee the potential consequences of developing a system and render them as measurable impacts.

8. **Assessors:** Impact assessments are conducted by assessors. The independence of assessors from the actor as well as the forum is crucial to how an assessment process identifies impacts, how those impacts relate to tangible harms, and how it acts as an accountability mechanism that avoids, minimizes, or mitigates such harms.
9. **Impacts:** These are proxies for harms produced through the deployment of a system in the real world. They enable the forum to identify and ameliorate potential harms, stipulate conditions for system operation, and thus, hold actors accountable.
10. **Harms and Redress:** Harms are lived experiences of the adverse consequences of a system's deployment and operation in the real world. Some of these harms can be anticipated through impact assessments as impacts, others cannot be foreseen. Redress procedures must be developed to complement any anticipated harm to secure justice.

While all ten of these components are necessary for robust accountability, policymakers do not need to fully specify all of them at the point of rule-making. Rather, policymakers should ensure that an AIA process is defined in a way that involves a diverse range of stakeholders, specifically including the communities impacted by algorithmic systems.

We see this demonstrated in impact assessments for other highly complex fields (such as environmental impacts). Policymakers are best positioned to hold space for industry, experts, and public interest advocates to refine and evolve specific methods over time that are responsive to the mandates developed via rule-making procedures and jurisprudence.

The most pressing step for policymakers is to establish the proper **source of legitimacy** for AIAs through legislation and/or policy. If this initial policy can specify **actors and forum, catalyzing event,** and the terms for **public consultation** and **access,** then stakeholders can collectively develop the remaining components over time.

A well-specified forum, that includes industry actors, community representatives, and independent advocates, ultimately will build consensus about the types of impacts that should be assessed and the steps needed to minimize or mitigate harmful impacts.

PUTTING PUBLIC INTEREST IN ALGORITHMIC ASSESSMENT

Our research highlights that establishing algorithmic accountability is both a matter of creating an obligation for developers to assess the impacts of algorithmic systems, and fostering an equitable, collaborative and deliberative process for creating standardized assessment practices. The latter has been neglected thus far in our analysis of legislation targeting algorithmic accountability.

- In 2019, Canada introduced an AIA requirement for any agency that procures AI systems. The legislation makes a brief risk-assessment checklist mandatory for

developers who vend through procurement processes for federal government agencies purchasing automated decision systems. This self-assessment lacks any obligation or **public consultation** with or study of affected communities.

- In early 2021, the European Union introduced a new regulatory framework that requires providers of high-risk AI systems to complete “conformity assessments” overseen by sector-specific regulators. The framework pays inadequate attention to how assessment methods will be assembled and vetted, and lacks opportunities for contestation over methods and **public consultation** and **redress** in response to assessment outcomes.
- While the US saw the introduction of the Algorithmic Accountability Act of 2019, the legislation did not pass and may be reintroduced during this Congress. The 2019 version of this legislation directed the Federal Trade Commission to require algorithmic impact assessments for certain high-risk automated decision systems. However, the bill’s definition of impact assessments does not consider how that process should incorporate important forms of expertise, thereby potentially excluding public interest advocates. It also does not require that AIAs be published or otherwise provide **public access.**

In each of these cases, the role of the public has been omitted or relegated to a “public comments period.” Unless policymakers enshrine a collaborative process that involves community members, public interest advocates and independent asses-

sors in the design of AIA assessment methods, **industry actors will have every incentive to turn assessment into mere checklist compliance.** Now is the moment to consider what changes to these assessment regimes will make them the most effective.

Algorithmic impact assessments are not only a means for measuring and preventing harm, but also integrating the interests and agency of affected individuals, communities, and the public writ large into those measurement practices. Impact assessment solely for compliance will not shift power or agency over systems to affected communities and leaves gaps in the resulting accountability regime.

Strengthening algorithmic impact assessments can create a stronger basis for people and communities to act as a forum that can exert influence over algorithmic systems and hold actors involved in developing and maintaining such systems accountable. The choices made about impact assessments, made clearer through the framework we develop in this report, determine whether these goals are achieved.

POLICY RECOMMENDATIONS

With this report, we have developed a framework to guide the collaborative, multi-stakeholder design of algorithmic impact assessments. Policy conversations around AIAs need to account for how stakeholders' expertise is brought to bear upon the crafting and execution of policy. Policymakers should also attend to the incorporation of community advocates into the rule-making process, to ensure that the assessment practices take into account the experience of being subject to algorithmic systems, and to protect the public

interest. This requires a process of "calling in": explicitly inviting stakeholders into the process and holding space for them to participate without being crowded out by other stakeholders who have historically held more power over the direction of policy.

Facilitating multi-stakeholder conversations about the methods of AIAs should be a priority. There is an existing field of both client-serving auditors and critical third-party auditors; however, these auditors have a limited set of tools to measure algorithmic impacts, and developers lack a regulatory incentive to collaborate and share findings. **Without a mandate to facilitate a consensus-driven, multi-stakeholder conversation about methods, we will see industry capture of methods for measuring and mitigating harms, rendering any attempted AIAs inadequate to the task of achieving accountability in the public interest.**

We recommend a series of meetings according to the themes and discussion questions outlined below. These meetings would bring together a wide range of stakeholders to collaborate on the design of an impact assessment process, and the development of qualitative methods to measure impacts as proxies for harms of algorithmic systems. This approach will build the expectation that algorithmic systems will be regulated in some way, create another mechanism through which policymakers can center public interest, and ensure that algorithmic systems will have socially and economically equitable benefits.

- **Defining how public interests are to be protected:** What process will ensure as exhaustive a set of algorithmic harms as possible will be considered as part of an assessment process? How will they be enumerated?

- Standardizing assessment methods:** What responsibilities should developers and vendors of algorithmic systems have to ensure that their systems are accessible to assessors? What documentation and access should they be expected to provide? What tools do communities and advocacy organizations already use to understand the algorithmic impacts with which they are concerned? What tools should assessors use to evaluate specific impacts? How can each set of stakeholders' expertise inform the others?
- Defining an accountability regime for algorithmic systems:** Under what circumstances should an AIA be required, and in what circumstances are they irrelevant? Who is responsible for overseeing and coordinating the assessment? Who is responsible for determining that an assessment is robust and adequate? What steps ought to be taken to minimize and mitigate impacts documented through assessment? Who ought to decide and enforce the execution of those steps?
- Comparing policy and regulatory tools and approaches:** What powers do regulators currently have? What do they need to build and enforce an accountability regime for algorithmic systems? How can we align and learn from state and local level policymakers, as well as international governments and organizations, who are already working on the regulation of algorithmic systems?

ANNEX

As an independent, nonprofit research institute, Data & Society focuses on the social and cultural issues arising from the increasing appropriation of data-driven technologies. We seek to inform and develop frames for discussion for these complex issues both through our own research and by supporting the voice of other stakeholders who face adverse consequences of these technologies on an everyday basis and are in the nascent process of articulating such consequences as matters of concern. We are a resource, a catalyst, and a convener as new sectors of society contend with the complexity of making data-driven technologies work and the need for trade-offs in increasing reliance on data as a tool for management and decision-making practices.

This research was conducted by the AI on the Ground Initiative (AIGI), a research initiative housed at the Data & Society Research Institute designed to address and act on existing knowledge gaps in the emerging terrain of algorithmic systems. The Initiative's goal is to understand and articulate issues of deploying algorithmic systems by conducting original research on the practices involved in the creation and lived experiences of such systems, and publishing research reports and academic articles based on this research. Based on a commitment to follow research with action, AIGI aims to develop new conceptual, legal, and design-based tools built from basic research outputs, in order to better engage and empower policymakers, scholars, and those building and affected by algorithmic systems.

If you'd like to learn more about our Algorithmic Impact Assessment policy recommendations, please contact **Brittany Smith**, Policy Director at policy@datasociety.net.