

Governing with Algorithmic Impact Assessments: Six Observations

By Emanuel Moss¹, Elizabeth Anne Watkins², Jacob Metcalf³, and Madeleine Clare Elish⁴

Abstract

Algorithmic impact assessments (AIA) are increasingly being proposed as a mechanism for algorithmic accountability. These assessments are seen as potentially useful for anticipating, avoiding, and mitigating the negative consequences of algorithmic decision-making systems (ADS). At the same time, what an AIA would entail remains under-specified. While promising, AIAs raise as many questions as they answer. Choices about the methods, scope, and purpose of impact assessments structure the possible governance outcomes. Decisions about what type of effects count as an impact, when impacts are assessed, whose interests are considered, who is invited to participate, who conducts the assessment, the public availability of the assessment, and what the outputs of the assessment might be all shape the forms of accountability that AIA proponents seek to encourage. These considerations remain open, and will determine whether and how AIAs can function as a viable governance mechanism in the broader algorithmic accountability toolkit, especially with regard to furthering the public interest. Because AIAs are still an incipient governance strategy, approaching them as social constructions that do not require a single or universal approach offers a chance to produce interventions that emerge from careful deliberation.

Introduction

From government policy makers to company board rooms, the idea of implementing “algorithmic impact assessments” (AIAs) as a form of algorithmic accountability is gaining momentum. These assessments are seen as potentially useful for anticipating, avoiding, and mitigating the negative consequences of algorithmic decision-making systems (ADS).⁵ Already, the EU has stipulated through its GDPR legislation that companies must provide privacy impact assessments in the interest of user rights.⁶ The Canadian government now requires a

¹ Data & Society Research Institute; CUNY Graduate Center (emanuel@datasociety.net)

² Data & Society Research Institute; Columbia University (ewatkins@datasociety.net)

³ Data & Society Research Institute (jake.metcalf@datasociety.net)

⁴ Data & Society Research Institute (mcelish@datasociety.net)

⁵ Andrew D. Selbst, “Disparate Impact in Big Data Policing,” 52 Georgia L. Rev. 109 (2017), <https://ssrn.com/abstract=2819182>; Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. “Algorithmic impact assessments: A practical framework for public agency accountability.” AI Now Institute. 1-22. <https://ainowinstitute.org/aiareport2018.pdf>.

⁶ Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of

checklist-style version of algorithmic impact assessments for its agencies that use algorithms.⁷ Companies like Facebook and Google are commissioning human rights impact assessments to identify harms of their platforms and products.⁸ The Algorithmic Accountability Act, proposed in the US Congress in 2019, would require companies with large user-bases to conduct impact assessments of automated decision systems that affect certain sensitive domains of people's lives.⁹

The term “algorithmic impact assessment” has been used as an umbrella term, referring to a range of processes and documentation, and emerges within the context of an expanding toolbox of potential accountability processes, including algorithmic audits, datasheets, “nutrition” labels, and model cards.¹⁰ The general idea of an AIA is to document the development and impact of an algorithmic system, providing a point of leverage for mitigating potential harms to individuals and vulnerable communities.¹¹ It's a compelling concept that leaves more questions than answers. What constitutes an assessment? An impact? An algorithm? An algorithmic system? Who gets to decide? Should algorithms used by private companies be subject to the same forms of accountability as those used by public institutions? What forms of accountability are at stake?

such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 (hereafter “GDPR”).

⁷ Government of Canada. 2019. Directive on Automated Decision-Making.

<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> See also

<https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>; Karlin, Michael and Noel Corriveau. 2018. “The Government of Canada’s Algorithmic Impact Assessment: Take Two.” Medium.com:

<https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f>.

⁸ Warofka, Alex. 2018. “An Independent Assessment of the Human Rights Impact of Facebook in Myanmar.” Facebook press release, Nov 5: <https://about.fb.com/news/2018/11/myanmar-hria/>;

Allison-Hope, Dunstan, Hannah Darnton, and Michaela Lee. 2019. “Google’s Human Rights by Design.” Business for Social Responsibility blog:

<https://www.bsr.org/en/our-insights/blog-view/google-human-rights-impact-assessment-celebrity-recognition>.

⁹ Office of Senator Cory Booker. 2019. Booker, Wyden, Clarke Introduce Bill Requiring Companies To Target Bias In Corporate Algorithms. Press release, Washington DC:

https://www.booker.senate.gov/?p=press_release&id=903

¹⁰ Raji et al. 2020. “Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing.” In Conference on Fairness, Accountability, and Transparency (FAT* ’20), January 27–30, 2020, Barcelona, Spain; Gebru et al. 2018. Datasheets for datasets. Technical Report. arXiv preprint arXiv:1803.09010; Holland et al. 2018. The dataset nutrition label: A framework to drive higher quality data standards. arXiv:1805.03677; Mitchell et al. 2019. Model cards for reporting model performance. In Proceedings of ACM Conference on Fairness, Accountability and Transparency (FAT*);

See also PAI’s About ML project: <https://www.partnershiponai.org/about-ml/>.

¹¹ The Ada Lovelace Institute and DataKind UK have pointed out that algorithm auditing (how does the system function and is it accurately described?) is often conflated with impact assessments despite algorithm auditing (especially bias auditing) being both more robustly fleshed-out and having a narrower purview. Ada Lovelace Institute and DataKind UK, 2020. Examining the Blackbox: Tools for assessing algorithmic systems.

<https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf>

Existing proposals for AIAs and related governance practices answer each of these questions differently. This is to be expected; there is not as of yet a clear coalescence of institutional, intellectual, regulatory, and judicial power around any particular vision of what an AIA *is*. In our opinion, this heterogeneity provides an important opportunity to critically shape the purpose and methods of AIAs going forward.

Algorithmic decision systems, including machine learning and AI techniques, present unique and substantial challenges when it comes to assessing their impact on society. These include, but are not limited to, how these systems are built, how they relate to the data used to train and retrain them, and the power relationships between agencies and industries that operate ADS, the complex role played by 3rd-party vendors, and how “users” and “the public” are constituted.¹² The existing body of research on how to audit, investigate, and understand undesirable and unexpected behaviors of such systems is currently growing, and is much needed.¹³ Moreover, there is a lack of empirical evidence and qualitative research to support how — or whether — algorithmic impact assessment will be an effective, or even a desirable, governance mechanism.¹⁴ A robust approach to assessing algorithmic impacts will couple these ongoing efforts with the capacity to consider the range of social, technological, and legal entities that are implicated throughout the process of developing an algorithmic system.

If the goal is to develop new and stronger mechanisms of accountability for the cascading effects of algorithmic systems, impact assessments offer many opportunities. Rather than relying on slow-moving legislatures to outline exactly what algorithmic systems can and cannot do, impact assessments — whether mandated or directly administered by a responsible government agency — can set standards for evaluating the performance of such systems and provide a means of accountability that tracks alongside the shift in power as it is transferred from lawmakers to agency personnel to those who perform impact assessments.¹⁵ Many would

¹² See for instance: Cath, Corrine. 2018. “Governing artificial intelligence: ethical, legal and technical opportunities and challenges.” *Phil. Trans. R. Soc. A* 376: 20180080.; Koene, Ansgar, Chris Clifton, Yohko Hatada, Helena Webb, and Rashida Richardson. 2019. “A governance framework for algorithmic accountability and transparency.” Brussels: European Parliamentary Research Service.; Veale, Michael and Brass, Irina, *Administration by Algorithm? Public Management Meets Public Sector Machine Learning*. 2019. In: *Algorithmic Regulation* (Karen Yeung and Martin Lodge eds., Oxford University Press, 2019). Available at SSRN: <https://ssrn.com/abstract=3375391>; Mulligan, Deirdre K. and Kenneth A. Bamberger. 2019. “Procurement As Policy: Administrative Process for Machine Learning”. *Berkeley Technology Law Journal*, Vol. 34, 2019. Available at SSRN: <https://ssrn.com/abstract=3464203>.

¹³ Raji et al., 2020..

¹⁴ Following the proposal of the Algorithmic Accountability Act, several organizations voiced their concerns about the legislative proposal, including one of the authors. See, New, Joshua. 2019. “How to Fix the Algorithmic Accountability Act.” Center for Data Innovation: <https://www.datainnovation.org/2019/09/how-to-fix-the-algorithmic-accountability-act/>; Barbanel, Jerry. 2019. “A look at the the proposed Algorithmic Accountability Act of 2019.” IAPP.org: <https://iapp.org/news/a/a-look-at-the-proposed-algorithmic-accountability-act-of-2019/>;; Selbst, Andrew, Madeleine Clare Elish and Mark Latonero. “Accountable Algorithmic Futures.” *Points*: <https://points.datasociety.net/building-empirical-research-into-the-future-of-algorithmic-accountability-act-d230183bb826>.

¹⁵ See Shapiro, David L. “The Choice of Rulemaking or Adjudication in the Development of Administrative Policy.” *Harvard Law Review* 78, no. 5 (1965): 921-72.; DeLong, James V. “Informal Rulemaking and the

argue that privacy impact assessments have been a valuable tool in a larger toolbox of privacy protecting processes.¹⁶ However, such assessments are not without their critics.¹⁷ Still, impact assessments have been productive in providing a basis for rational decision-making between competing alternatives in the design of a development project, where tradeoffs between potential upside benefits and downside impacts must be made.¹⁸

At the same time, the efficacy of impact assessments has been critiqued in the context of data-driven systems, as well as in other domains, including fiscal impact assessments and environmental impact assessments.¹⁹ These critiques have mainly focused on the role that impact assessments play in creating systems of superficial self-regulation or the mere veneer of accountability.²⁰ A valuable area of study will be analyzing the methods and lessons of parallel domains, including work currently being undertaken by the authors. As a first step, in this paper we lay out observations about the range of practices and forms of accountability that circulate across various types of impact assessment. In forthcoming work, we analyze a collection of case studies expanding on these observations with the goal of generating recommendations. This research will allow us to ask: to what extent will algorithmic impact assessments require the creation of new definitions and processes, and to what extent are existing processes feasible or

Integration of Law and Policy." *Virginia Law Review* 65, no. 2 (1979): 257-356.; West, William.

"Administrative Rulemaking: An Old and Emerging Literature." *Public Administration Review* 65, no. 6 (2005): 655-68.;

¹⁶ Waldman, Ari Ezra, "Privacy Law's False Promise". 2019. *Washington University Law Review*, Vol. 97, No. 2, 2020. Available at SSRN: <https://ssrn.com/abstract=3499913>.

¹⁷ Wright, D., 2011. Should privacy impact assessments be mandatory?. *Communications of the ACM*, 54(8), pp.121-131; Wadhwa, K. and Rodrigues, R., 2013. Evaluating privacy impact assessments. *Innovation: The European Journal of Social Science Research*, 26(1-2), pp.161-180. Wright, D. and Friedewald, M., 2013. Integrating privacy and ethical impact assessments. *Science and Public Policy*, 40(6), pp.755-766. For critiques, see Bamberger, Kenneth A, and Deirdre K Mulligan. 2008. "Privacy Decisionmaking in Administrative Agencies." *The University of Chicago Law Review*, 34.;

¹⁸ Steinemann, Anne. 2001. "Improving Alternatives for Environmental Impact Assessment." *Environmental Impact Assessment Review* 21 (1): 3–21.

¹⁹ Kaminski, Margot E. and Gianclaudio Malgieri. 2019. "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations". U of Colorado Law Legal Studies Research Paper No. 19-28. Available at SSRN: <https://ssrn.com/abstract=3456224>; Mauer, Marc. 2007. "Racial Impact Statements as a Means of Reducing Unwarranted Sentencing Disparities." *Ohio State Journal of Criminal Law* 5: 28. Also, see, Taylor, Serge. 1983 (*Making Bureaucracies Think: The Environmental Impact Statement Strategy of Administrative Reform*. Stanford, CA: Stanford University Press) for a canonical study of how the NEPA environmental impact assessment process facilitates development projects despite foundational intentions to balance competing interests of environmentalists and federal agencies. See also Goldman, Michael. 2005 (*Imperial Nature: The World Bank and Struggles for Social Justice in the Age of Globalization*. Yale Agrarian Studies Series. New Haven, CT: Yale University Press) for a study of the World Bank's use of an environmental impact assessment process to depoliticize international development and "greenwash" exploitative economic development projects in the developing (ie, post-colonial) world.

²⁰ Waldman, Ari Ezra. 2020. "Cognitive Biases, Dark Patterns, and the 'Privacy Paradox.'" *Current Opinion in Psychology* 31: 105–9.; Mourey, James A. and Ari Ezra Waldman. 2020. "Past the Privacy Paradox: The Importance of Privacy Changes as a Function of Control and Complexity". *Journal of the Association for Consumer Research* 5:2, 162-180.

desirable? Would an algorithmic impact assessment *supplement* or *replace* existing impact assessments already on the books?

Can algorithmic impact assessments be effective governance and accountability mechanisms for algorithmic and data-driven sociotechnical systems, and if so, how? We need more research and analysis before drawing conclusions. As a contribution to this growing area of inquiry and action, we draw on our backgrounds in the social sciences and our experience studying and analyzing the consequences of automated and AI technologies in order to think through the recent history of impact assessment, and what lessons might be learned for *algorithmic* impact assessments. We do so to identify how an algorithmic impact assessment process might reasonably reduce harms to individuals and groups and minimize disruptive impacts of algorithmic systems, while still producing useful and beneficial algorithmic technologies. To this end, we offer six observations that must be grappled with when assessing impact assessments:

- What constitutes an impact is non-obvious.
- Different types of impact come into focus depending on when an assessment occurs.
- Public participation in an assessment process is not synonymous with accountability to the public.
- Impact assessments structure how institutions operate and interact.
- Assessing impacts does not necessarily mean addressing harms.
- Impact assessments ask us how the world might be otherwise.

Six Observations

What constitutes an impact is non-obvious.

There is no pre-existing or universal definition of an “impact” that can be applied in the context of an impact assessment because there is a central confounding question about how to bound the impact to be assessed. “Impact” invokes a causal relationship: an action taken by an organization (or a system an organization operates) *causes* an *effect* in the world and thereby *impacting* some aspect of the world by making it otherwise. However, it is difficult to delineate a clear causal relationship for most phenomena we are interested in measuring as impacts. This inevitably raises the question: what can be identified as an impact resulting from one particular cause, and how can that cause be properly identified as having stemmed from something that the organization has control over?

The process of identifying, measuring, formalizing and accounting for “impacts” is a power-laden process that does not have a neutral endpoint. Because these systems are complex and multi-causal, defining what counts as an impact becomes a domain of contestation, shaped by social, economic and political power. Which impacts get assessed is ultimately the result of decisions to include an impact as assess-able. The list of impacts considered assess-able will necessarily be incomplete, and assessments will always be partial.

For most existing types of impact assessment, the domain of impact is relatively well-bounded and the assessment examines the impact of an undertaking *to* a particular resource, domain, or right. For instance, privacy impact assessments examine impacts *to* privacy, human rights impact assessments examine impacts *to* human rights and environmental impact assessments assess impacts *to* the environment.

In contrast, how does one bound the impacts *of* algorithmic systems within an impact assessment process? The domains of ADS's are expansive and the domain of any given AIA could be similarly broad — one need look no further than credit scores to sense just how expansive the set of impacts from an ADS might be. What types of impacts are reasonable to attribute to algorithmic systems, and not to other causes? Similarly, where an effect is multi-causal how can an assessment attribute a reasonable degree of responsibility to those implementing an algorithmic system? Impact may arise from the data used to train the model, from the algorithmic techniques and design specifications employed in the model, or from the context in which it is applied in the real world. Importantly, the components of an algorithmic system may be assembled from many different sources of data, using many different open and proprietary code bases, and be used in manners quite tangential to their original purposes. Many different parts of a company and/or users of a system may be implicated by different components. In thinking through algorithmic impact assessments as a new type of impact assessment, answering questions about what counts as an impact, and how those impacts might be measured and used for any sort of rational evaluation, is crucial.

Different types of impact come into focus depending on when an assessment occurs.

Critical to unpacking *what* we mean by “impact” is a consideration of *whether* impacts can be exhaustively enumerated, and how the capacity to document impacts depends on *when* we assess that impact. One of the core questions is whether an impact is assessed *ex ante* (before the event) or *ex post* (after the event).

If impacts are assessed *ex ante*, the assessment is a prediction about the risks and consequences of a proposed system. Generally speaking, *ex ante* assessments are based on existing information like prior use cases, empirical measurements of the behavior of the system in testing environments, or narrative records of how the system was designed and iteratively developed.²¹ Environmental impact reports, data protection impact assessments, and fiscal impact assessments are based on *ex ante* assessments.

In contrast, if impacts are assessed *ex post*, the assessment is a record of information that is primarily gained following the introduction of a product or system. Generally speaking, this

²¹ The prior knowledge necessary to anticipate, measure, and mitigate impacts is not without contention, as even baseline data about specific environmental quality measurements can be manipulated through the environmental impact assessment process. See Kinchy, Abby. 2020. “Contentious Baseline: The Politics of ‘Pre-Drilling’ Environmental Measures in Shale Gas Territory.” *Environment and Planning E: Nature and Space* 3 (1): 76–94.

information might include field observations, interviews with stakeholders, or measurements of system outcomes in real, rather than simulated, environments. Examples of impact assessments that use *ex post* assessments are supply chain assessments and human rights impact assessments.

Proponents of *ex ante* methods might argue that this approach creates invaluable opportunities to assess a project and accordingly modify design prior to its release. In the case of environmental impact assessments, for example, the public debates that occur *before* a development can begin are critical spaces to voice dissent. Proponents of *ex post* approaches might argue that it is the impacts that we are least equipped to predict that are the ones that are likely to be most important to observe and assess. These impacts are likely only be knowable *post facto*, when a system has been deployed and integrated in specific social contexts.

When and how an impact is assessed not only affects the types of impacts that can be assessed, but also the kinds of processes that need to be established for an assessment to lead to organizational accountability. *Ex ante* assessments ask what the anticipated impact of decisions might likely be, while *ex post* assessments ask what would have happened had a different choice been made, and, by implication, what would happen if a different choice is made going forward. These two forms rest on differing theories of change, meet different organizational demands, and posit different relationships between cause and effect.

They differ in how they view, frame, and describe choices and impact. While *ex post* analyses imagine how an agency or business might intervene in an ongoing process, *ex ante* analyses ask assessors to imagine the potential rewards or risks at stake, and must bracket away the difficulties of anticipating the outcomes of interventions in complex sociotechnical systems.²² Although the approaches can be complementary over the life cycle of a system, assessments are temporally bounded, and there are tradeoffs involved in choosing one approach over the other.

The distinction between *ex ante* and *ex post* assessments demonstrate that different types of impacts come into focus at various moments in any impact assessment process, and that the impacts of a system can only be artificially bounded. The impacts that are discernible at the design and specification phase of a project are different than the impacts that become visible in other phases, particularly for information technologies that continuously patch, update, and scale. For that reason, *ex ante* assessments may be most useful as a form of transparency for technical or historical records. Particularly for algorithmic systems, which may need to be assessed in terms of how interpretable or explainable their outputs are to human users, having a record of choices made in design (*ex ante*) are prerequisites for any forensic (*ex post*)

²² Bailey, P.D., Haq, G. and Gouldson, A., 2002. "Mind the gap! Comparing *ex ante* and *ex post* assessments of the costs of complying with environmental regulation." *European Environment*, 12(5), pp.245-256.

investigation.²³ And to fully understand impacts that produce harm to people requires careful consideration about when it becomes possible to anticipate, detect, and mitigate such harms.

Impact assessments structure interactions in and between institutions.

Impact assessments bring different sets of organizations into relation with each other, through formal structures specified by that type of impact assessment. These relationships have economic and political consequences. For instance, the structures and relationships that are established also set the conditions for different types of accountability. Some impact assessment regimes establish a public process by which different (often antagonistic) actors, including the general public, are formally brought into dialogue. Other impact assessment regimes necessitate ongoing interaction between actors in ways that establish more collaborative rather than antagonistic modes of interaction. Assessments statutes create frameworks within which policymakers and technical actors are constrained and empowered when it comes to the design and implementation of a particular instrument.²⁴

According to Serge Taylor's analysis of the first ten years of the Environmental Impact Assessment (EIA) process called for under the 1969 National Environmental Protection Act (NEPA), the EIA process places environmental advocacy organizations into an adversarial relationship with project developers, which plays out through formal bureaucratic procedures within the Environmental Protection Agency (EPA).²⁵ The EIA process places developers, bureaucrats, environmental analysts, and advocacy organizations into a specific set of relations by requiring a proposed development plan be assessed by experts according to established guidelines before a project can move forward. The human rights impact assessment (HRIA) process, however, places institutions into very different relationships with each other. Human rights experts are contracted by a corporate entity to produce an analysis of their business activities, and that analysis serves as a knowledge base from which that corporate actor may make voluntary choices to address potential human rights impacts within their control.

Different regimes of impact assessment, therefore, evoke specific forms of social and political power — between bureaucrats, developers, and public advocates, or between businesses and those whose human rights are impacted by business activities — that must be properly interrogated to properly scope a new impact assessment process. Nevertheless, over time an impact assessment regime can shift as new actors (agency departments, consulting companies, professional roles) respond to the demand for the work needed to complete impact assessments. New economies of compliance are created, and new entities can arise to take on duties that were intended to be performed by others, as with environmental consulting firms for the EIA process.²⁶ As decision-making power shifts, so too does the locus of power, and

²³ Selbst, Andrew D. and Barocas, Solon, 2018. "The Intuitive Appeal of Explainable Machines." 87 *Fordham Law Review* 1085 (2018).

²⁴ Solow-Niederman, Alicia, YooJung Choi, and Guy Van den Broeck. 2019. "Institutional Life of Algorithmic Risk Assessments." *Berkeley Technology Law Journal* 34 (705): 05–744.

²⁵ Taylor, 1983.

²⁶ Waldman, 2019.

accountability for that power must also shift accordingly. And much of the meaning and intent of compliance can shift toward terms favorable to powerful actors (and perhaps contrary to the interests of the the public good), as when firms gradually take on the role of compliance for themselves and are only required to attest to their own compliance to an agency that is in a position of oversight.²⁷

In designing a new impact assessment process, particularly for AIAs, how relationships between organizations are structured is an important point of leverage that ought to be the subject of deliberation before formalizing an impact assessment as a regulatory requirement. The existing technological development process already typically involves documentation, and adding impact assessment-related specifications to existing documentation processes could be minimally disruptive, although this might differ between startups and more mature organizations. Understanding how relationships between organizations have been structured by other types of impact assessment and documentation processes will be crucial for deliberating over new types of impact assessment.

Public participation in an assessment process is not synonymous with accountability to the public.

Public participation is a critical component of democratic governance. In federal agency rule-making it is a key mechanism for making the government more responsive and accountable to the public.²⁸ In environmental decision making, for example, public participation plays a strong role in education and resolving issues of conflict and mistrust.²⁹ The commenting process, further, can change an agency's course of action.³⁰ Additionally, the legitimacy of the impact assessment process depends on some degree of participation from a variety of stakeholders, including government agencies, private companies, consulting firms, and advocacy groups, as well as some definition of "the public".³¹

However, not all forms of participation are equal. Different types of impact assessment mobilize different forms of representation and participation from respective constituencies through public comment periods, focus groups, rapid assessments, or public meetings.³² How participation is

²⁷ Edelman, Lauren B. et al. 2011.

²⁸ Rowe, G. and Frewer, L.J., 2000. Public participation methods: A framework for evaluation. *Science, technology, & human values*, 25(1), pp.3-29.

²⁹ Beierle, T.C. and Cayford, J, 2010. *Democracy in practice: Public participation in environmental decisions*. Routledge.

³⁰ Kochan, D.J., 2017. The commenting power: Agency accountability through public participation. *Okla. L. Rev.*, 70, p.601.

³¹ See Jonathan Poisner, A Civic Republican Perspective on the National Environmental Policy Act's Process for Citizen Participation, 26 ENVTL. L. 53, 55 (1996) ("[C]itizen participation in the creation of NEPA-mandated [EISs] has, in all likelihood, spawned the largest amount of citizen participation in environmental decision making over the last two decades.").

³² For an introduction, see Fung, Archon. 2015. "Putting the Public Back into Governance: The Challenges of Citizen Participation and Its Future." *Public Administration Review* 75(4), pp.513-522; and Involve. 2005. *People & Participation: How to put citizens at the heart of decision-making*. Beacon Press:

measured, and how “successful” participation is defined, are debated topics.³³ Efforts to bring together community stakeholders can create new spaces of deliberation and empowerment.³⁴ However, without careful planning, they can also inadvertently flatten asymmetries in agency, power, voice, and vulnerability. Scholarship in stakeholder theory, for example, finds that “stakeholders” are identified differently between institutions, with disparate definitions contingent on one’s power to influence the firm, the legitimacy of one’s relationship with the firm, and the urgency of one’s claim.³⁵ Moreover, despite the best of intentions, the relationship between public participation, transparency, and accountability, is far from straight-forward.³⁶ While critically important to a functioning and accountable democracy, public participation is not a panacea for the potential negative impacts of ADS. Poorly protected commenting procedures can be easily gamed by actors seeking to discredit their validity.³⁷ A lack of rigor and reflexivity about participation processes risk them becoming a performance of caring for those who might be impacted, or even enrolling vulnerable populations into harmful processes, or making their vulnerability legible to bad actors.

Assessing impacts does not necessarily mean addressing harms.

An impact assessment itself does nothing to mitigate or directly address identified harms, although some assessment processes require mitigation of impacts to be explicitly documented. Rather, impact assessments provide information upon which other interventions or processes can build. Without identifying what impacts are, or what they are likely to be, it is impossible to mitigate harmful impacts, or govern a response to those impacts-- and ultimately, to hold responsible parties accountable for those impacts. For most extant impact assessment processes, a great deal of attention has been paid to methodologies that can provide a knowledge base on which properly empowered actors can engage in rational decision-making.³⁸ What constitutes a rationale for decision-making in the context of particular impact assessment regimes is an extension of how a particular form of impact assessment is imagined to facilitate

London. Available online:

<https://www.involve.org.uk/sites/default/files/uploads/People-and-Participation.pdf> .

³³ Rosener, J. B. 1978. “Citizen participation: Can we measure its effectiveness?” *Public Administration Review*, September/October, 457-63

³⁴ Young, Meg, Lassana Magassa, and Batya Friedman. 2019. “Toward inclusive tech policy design: a method for underrepresented voices to strengthen tech policy documents.” *Ethics and Information Technology* 21: 89-103; Costanza-Chock, Sasha. 2020. *Design Justice: Community-led Practices to Build the Worlds We Need*. MIT Press: Cambridge, MA. Martin Jr., D., Prabhakaran, V., Kuhlberg, J., Smart, A., & Issac, W. S. 2020. *Participatory Problem Formulation for Fairer Machine Learning Through Community Based System Dynamics*. Fair & Responsible AI Workshop. CHI 2020.

³⁵ Mitchell, R.K., Agle, B.R. and Wood, D.J., 1997. “Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts.” *Academy of management review*, 22(4), pp.853-886.

³⁶ See for instance, Fox, Jonathan. 2007. “The uncertain relationship between transparency and accountability.” *Development in Practice*, 17:4, 683-671.

³⁷ Grimaldo, J. 2018. “New York Attorney General’s Probe Into Fake FCC Comments Deepens.” *Wall Street Journal*.

³⁸ See <https://www.iaia.org/key-citations.php>, especially Blakly, Jill, and Jessica Russell. 2018. “Trends, Issues and Insights in Cumulative Effects Assessment: A Review of International Academic Literature 2008-2018.” International Association for Impact Assessment (IAIA).

further decision-making. Human rights impact assessments imagine corporate actors as willing to make changes to their business practices following an assessment and furthermore provide a mechanism for remedy for individuals who have been harmed. EIAs imagine impacts to environmental resources can be anticipated in advance of a development project so that less impactful design choices can be made or mitigation efforts can be mandated.

In order to satisfy the sometimes-competing goals of developers, government agencies, and advocates to undertake projects while limiting harmful impacts, there are necessarily trade-offs between known impacts and the overall benefits of an undertaking to society. Understanding how to assess the scope, scale, and depth of an impact will be necessary for understanding when a potential impact is acceptable, within the context of a given project. And understanding when particular outcomes ought to be addressed by altering or abandoning a project, and how to go about addressing those needed changes, is crucial for the outcomes of any impact assessment process to fully realize goals of accountable governance that centers those most likely to be impacted by development projects.

For ADS, understanding how to assess potential harmful impacts and evaluate those harms against the potential benefits of the ADS will be crucial. Similarly to other forms of impact assessment, any rigorous AIA will likely detect harms that go unremedied, but the overall process should be able to facilitate robust, engaged, and transparent decision-making around what the tradeoffs are between potential harms and likely benefits.

Impact assessments ask us how the world might otherwise be.

Impact assessments, by drawing attention to design choices and consequences, prompt a consideration of alternatives. By creating room for such alternatives in policy or product development cycles, impact assessments can shape bureaucratic or corporate decision-making, potentially leading to different and more thoughtful design choices.

Impact assessments create an opportunity to reorganize power. Impact assessments have the potential to provide a lever of influence for figures who may not otherwise hold power to shape policy. Impact assessments might be a means for such communities to highlight otherwise overlooked or unforeseen sets of causes and effects.³⁹ This is the approach that motivated the creation of the National Environmental Protection Act. Drafters of that legislation chose not to pursue traditional methods of reform, like introducing additional legislation or engaging in drawn out political battles over agency leadership, and instead hoped that requiring agencies to make their environmental impacts transparent to the public would produce changes in policy.⁴⁰

³⁹ “What is Impact Assessment?” 2014. Based on OECD Directorate for Science, Technology and Innovation, “Assessing the Impact of State Interventions in Research – Techniques, Issues and Solutions”, unpublished manuscript.

⁴⁰ Taylor, 1984.

The models at the core of an algorithmic system are functionally proposals for how the world ought to be, made concrete through deployment and integration of that system with existing sociotechnical systems. They promise, for example, more efficient allocation of state resources, better rates of disease diagnosis, and optimized traffic flows. Relatedly, impact assessments, propose to make visible how the world might be changed by a specific project — algorithmic or otherwise.

Conclusion

As policy makers and industry actors develop AI governance, it is crucial to proceed carefully. Every governance structure will have benefits and drawbacks, and the devil is often in the proverbial details. As we have outlined in these challenges, impact assessments encompass a wide range of approaches, methodologies, and opportunities. There is no universal path to follow. These challenges also point us towards the need for empirical research and social-science methodologies to better inform that which assessments are intended to assess, as well as how assessment practices intersect with other social processes in particular contexts, from economic development to the administration of justice to the cultural significance of demographic categories. Doing so would extend the original intention of an impact assessment regime to introduce grounded, empirical science to policy decision making.⁴¹

Impact assessments of *algorithmic systems* in particular hold many challenges. On one hand, the algorithmic development process already presents several steps that could serve as ready-made handles for an impact assessment process to grab onto. Data collection design, data cleaning, model evaluation, and model deployment all represent moments when metadata relevant to the potential impacts of an algorithmic system can be documented for the assessment process. These are also moments when interventions might be made to mitigate any potentially harmful impacts prior to deployment. On the other hand, there remains a great deal of ambiguity around how impacts are defined (and by whom), how they are assessed (and by whom), and how this establishes or fails to establish robust forms of accountability.

This brings us to an important point: impact assessments are social constructions. They delimit certain systems as capable of causing impacts, stabilize abstract concepts (like rights, environmental resources, or privacy) as tangible entities capable of being impacted in concrete and measurable ways. They are material instantiations of social and political priorities about what is worthy of assessment and what impacts are tolerable. Recognizing the socially constructed character of impact assessment early in the development of a new assessment process offers a chance to produce interventions that emerge from careful, politically engaged deliberation.

Acknowledgements: This working paper is part of a larger project within the AI on the Ground Initiative at Data & Society examining algorithmic impact assessments. We are grateful to our collaborators on this project past and present, including Andrew Selbst, Mark Latonero, and

⁴¹ Taylor, 1984.

Aaina Argarwal. Our thinking has also benefited from ongoing conversations with Janet Haven, Jenny Brennan, Patrick Davidson, Sareeta Amrute, Hannah Darnton, Michelle Gilman, Amanda Lenhart, Robyn Caplan, danah boyd, Inioluwa Deborah Raji, and participants in the Data & Society RAW reading series.