Policy Memos

Canadian Commission on Democratic Expression

Learning Session 5: How can regulators and publics determine the risks, harms and impacts of algorithmic decision-making systems?

Thursday, Nov. 18, 2021 | 1:00 p.m. – 2:30 p.m. ET (UTC -4:00)

Virtual event via Zoom

Abstract of session

From healthcare, education, housing and law enforcement, algorithmic decision-making systems are being deployed in public and private sectors around the world, often without meaningful consent, privacy protection, or due diligence for those who stand to be most affected by their decisions. Currently, regulators and lawmakers are able to respond to such risks and harms caused by algorithmic and automated systems only after they have already happened. Often, these responses exclude the perspectives, voices, and ways of knowing of historically marginalized communities.

Information about the potential risks and impacts of these systems is needed to ensure policymakers and regulators have adequate information to determine the trade-offs and collective impacts of high-stakes systems. Discussions about and actions taken to address the potential risks and impacts of these systems must engage with how different groups experience, use, and relate to technologies. Meaningful engagement with Indigenous communities and epistemologies can afford us a different set of tools to think through automated decision making and ethical relations to technologies more broadly.

Policy questions:

Are current transparency mechanisms sufficient for reporting on algorithmic decision-making systems?

Should governments adopt public registers of algorithmic decision-making systems? If so, how should they be structured and governed?

How can policymakers build meaningful relationships with Indigenous communities to better address the risks, harms, and impacts of automated decision-making systems and digital technologies more broadly?

A true democracy promotes accessibility for everyone and is representative of the diverse and rich Indigenous Epistemologies that exist across Canada.

The Problem -- Historically Indigenous knowledge was not recognized or acknowledged by colonial settlers who were intent on maintaining the façade of their racial superiority for the purposes of occupying and taking Indigenous land. As a result, mainstream educators, policymakers and government bodies need to be educated on the importance of Indigenous knowledge and what it constitutes, as they currently have no reference point for it and remain unaware.

What is Indigenous Knowledge? – Indigenous knowledge are the teachings that are passed down from Elders, Traditional Teachers and Knowledge Keepers. Some examples of Indigenous knowledge teachings are; the Medicine Wheel, the Seven Grandfather Teachings, the story of the Peacemaker, etc. There are thousands of teachings across Canada that speak to the rich philosophies, epistemologies and worldviews held by First Nations, Metis and Inuit Nations.

Issues of Accessibility – Ignorance of Indigenous knowledge has created a funding gap in Canada for education curriculum development. For example, Ministries of Education, each have their own provincial budget and develop curriculum independently with their appointed and approved educators. Many of these curriculum developers do not have relationships with local Indigenous communities, which means they do not have access to the people who can facilitate and share Indigenous Knowledge Teachings.

The Fix – Indigenous Knowledge Education Online. Indigenous people see the Internet and New Technology as a potential space to address the oppression, racism, and ignorance that has been perpetuated by colonialism. As Justice Murray Sinclair noted, 'Education got us into this mess, and education will get us out.' My research of Indigenous Knowledge online has revealed that online spaces can be defined and validated through cultural protocols. These particular spaces are called *Digital Bundles* because they validate and center the Indigenous Knowledge Teachings of Elders, Traditional Teachers and Knowledge Keepers (Wemigwans,2018).

Distinct from the field of Indigenous Studies and digital storytelling and Indigenous new media art -- digital bundles are recognized and accepted as Indigenous Knowledge Education because they are community based and validated according to diverse cultural protocols.

My research has demonstrated that access to more Indigenous Knowledge Education online would not only increase the viability of such knowledge but would facilitate better relationships between settlers and Indigenous communities. For example, Larry Chartrand, a Metis Law Professor from Ottawa said,

> I would like to see perhaps more in-depth sites of relevant legends and stories of particular Nations, whether it is the Dene or the Coast Salish, because one of the most important aspects of a viable legal tradition is to ensure its accessibility to those who are expected to be bound by it.

A corrections officer from the North West Territories states,

> Further, Canadians are not educated on what has happened to Aboriginal people. We need to have more sites online that teach people about the treaties and about history from an Aboriginal perspective. We need these resources because they are not offered in the schools. These historical truths are not offered anywhere.

Digital Bundles complement the work of Sandy Grande's Red Pedagogy, where the foundation, perspective and production remain distinctive, rooted in Indigenous Knowledge and praxis (2004, 81).

Policy Recommendation – A funding body/ Centre that is dedicated to supporting the production of Indigenous Knowledge digital projects for the purposes of life-long learning for Canadians and Indigenous people. Significant knowledge of history and treaties, tactics of colonization and assimilation, and the ongoing quest to live with dignity as Indigenous People in Canada can all be conveyed to the broader public using the Internet and digital technologies.

Democratic Expression via Indigenous Knowledge Education is the answer to online hate, disinformation and ongoing racism, specifically with respect to Indigenous communities, who represent some of Canada's most vulnerable communities.

Indigenous Knowledge Education online represents an active opportunity for relationship building and a cost-effective way for addressing many of the Truth and Reconciliation Commission's Calls to Action.

Sources

Grande, S. (2004). Red Pedagogy: Native American Social and Political Thought. Lanham, MD: Rowman and Littlefield.

Wemigwans, J. (2018). A Digital Bundle: Protecting and Promoting Indigenous Knowledge Online. Regina, Saskatchewan: University of Regina Press.

Truth and Reconciliation Commission of Canada, (2015). Honouring the truth, reconciling for the future: summary of the final report of the Truth and Reconciliation Commission of Canada. https://publications.gc.ca/site/eng/9.800288/publication.html

Bio

Jennifer Wemigwans is an Assistant Professor in the Adult Education & Community Development program at OISE University of Toronto, where she teaches Indigenous Knowledge and Research Methods and pursues research on Digital Bundles through multiple technologies such as interactive documentary, virtual reality, augmented reality, etc. She also consults directly with Indigenous communities who are finding innovative ways to strengthen community through the use of online platforms.

Briefing to the Canadian Commission on Democratic Expression on Algorithmic Accountability on Online Platforms

Divij Joshi[*]

Dear Commissioners,

Thank you for inviting me to take part in this important discussion on the governance of online platforms.

Algorithms are important determinants of our online media and information environment. They are used in the moderation of online speech – to prioritise how and what information is conveyed, including to censor particular kinds of speech and privilege others. They are used to surveil, profile, categorise and target individuals and groups online. They do this in ways that often result in harms to human rights and human dignity, and in ways that have broader social consequences for values like fairness, transparency, accountability and democratic participation.

In 2021, I conducted research on a project assessing 'algorithmic accountability' for the public sector,[1] reviewing and analysing the implementation of policies intended address the harms caused through using algorithmic systems within the context of public administration. Below, I offer some of the lessons we learned from this review of governance mechanisms for algorithmic systems, which may have significance for the governance of online platforms.

Risk-Based Approaches to Algorithmic Governance

- Identifying algorithmic harms and risks is an important factor in determining the appropriate scope of governance and regulatory measures. Many jurisdictions have proposed or adopted 'risk-based' approaches towards the governance of algorithms, which ensure scalable and proportionate responses to regulatory compliance. Assessments of risk and regulatory scope have been based on the nature of the technology (eg., focusing on machine learning systems), the decision-making

[*] Doctoral Researcher, UCL Faculty of Laws.

[1] 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.

process (eg. the availability of human intervention or oversight in the final decision), as well as the context in which an algorithmic system is used, and the rights and interests it is likely to affect.

- While risk-based approaches to governance can introduce efficiencies, reduce compliance costs, and contribute to the effective use of limited resources for regulatory oversight, risk-based approaches are fundamentally limited, in that they presuppose quantifying and modelling risk of harms from the use of algorithmic systems, which can ignore or downplay complex and diffuse harms (Cohen, 2016).

- Measuring 'risk' in the context of online algorithms is not a straightforward task for even the most well-resourced researcher or regulatory agency. Algorithmic systems are complex, dynamic and networked systems, the effects of which are difficult to model in order to establish causal relationships between the algorithm and a material 'risk' or harm (Seaver, 2019). Algorithmic harms may also be diffuse, cumulative, and difficult to clearly quantify and measure. An alternative to such risk-based regulation, the precautionary approach, interrogates more critically the assumptions and knowledge that go into the determination or measurement of risk, and is based on preventing (rather than managing) unforeseen risks (Gellert, 2015). **As such, governments should also consider adopting uncertainty-based approaches in certain contexts, including precautionary approaches which require caution in the face of risks which have not been quantified under prevailing risk-measurement methodologies.**[2]

Algorithmic Impact Assessments

- There are few standards, or even shared vocabularies for clearly identifying and addressing what interests might be affected by algorithmic systems, particularly *ex ante,* as risk measurement and modelling often requires. It is, therefore, necessary to develop methodologies for the identification of risk and harm in algorithmic systems, suitable to different contexts, as a precursor to risk-based regulation.

[2] Such approaches are, to an extent, evident in legal restrictions on the use of facial recognition technology in various contexts globally.

- Algorithmic Impact Assessments ("AIAs") refer to an emerging category of governance processes which attempt to document and describe the potential risks or the 'impacts' of an algorithmic system used in particular contexts, in order to mitigate harms before they occur. Regulatory or policy mechanisms to document, quantify and mitigate risks of algorithmic systems have been proposed or implemented for the public sector in Canada,[3] New Zealand,[4] and Mexico[5]; for social media in India, [6] UK,[7] and for sectoral product safety in the EU.[8] AIAs draw, to varying degrees, from impact assessment methodologies for studying networked information systems, including Human Rights Impact Assessments, Privacy Impact Assessments, and Data Protection Impact Assessments, as well as a longer history of impact assessments in domains of environment, regulation, etc.

- AIAs can make visible the subjective choices of the actors (in this case, the online platforms) responsible for the use of an algorithmic system, including how they define harms, benefits and risks,

---

[3] 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

[4] 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] Government of India. (2021) The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 4(4), Available at: https://prsindia.org/files/bills_acts/bills_parliament/2021/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf

[6] Government of Mexico, (2018). Impact Analysis Guide for the development and use of systems based on artificial intelligence in the public sector, Available at: https://www.gob.mx/innovamx/articulos/guia-de-analisis-de-impacto-para-el-desarrollo-y-uso-de-sistemas-basadas-en-inteligencia-artificial-en-la-apf.

[7] Government of UK, Draft Online Safety Bill. (2021). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf.

[8] European Commission. (2021). Proposal For A Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

how they design measures for risk management and mitigation, and how they assess and choose between alternative options related to the deployment of an algorithmic system. They produce documentation and provide information to a range of actors (which may include the party deploying the algorithmic system, regulatory bodies, communities affected by the system, or civil society and the public) who can then use these frameworks to assess the credibility and appropriateness of an algorithmic system (Metcalf et. al., 2021).

- There is not enough evidence to conclude that AIAs are an effective mechanism for ensuring accountability for harms caused by online platforms through the use of algorithms. However, our research provides insights into the factors that influence whether AIAs achieves its intended objectives:

    - 'Impacts' under AIAs are constructs which are intended to map to material harms experienced by people. Impacts must be constructed in a way which both render harms quantifiable, as well as actionable by regulators or regulated parties. Ultimately, much of the effectiveness of AIAs rests on the ability of the IA process to accurately translate material harms faced by persons subjected to algorithmic systems into actionable 'impacts', which brings into focus the values and epistemic boundaries considered in the impact assessment processes. **Mapping the impacts of algorithmic systems to real-world harms requires the meaningful participation of the groups and individuals, or the broader public affected by algorithmic systems in the process of conducting an AIA. The process of an AIA must also assemble the diverse forms of expertise in systematically studying algorithmic systems and harms, including researchers of algorithmic systems, social science, law and governance, as well as contextually-situated expertise from affected communities to identify relevant values and forms of knowledge appropriate to mapping actual experiences of harm to actionable 'impacts'.**

    - In some cases, the lack of transparency of the IA process and outcomes to a critical audience (the general public, or civil society) has been criticised as a structural impediment to accountability. **The process of constructing, documenting algorithmic impacts, along with the choices made in response to documented harms must be transparent, with the documentation about AIAs made transparent to the public in a meaningful manner. This involves not only releasing**

documentation, but appropriately engaging diverse and critical audiences to understand and respond to the AIA process and outcomes.

- AIAs may be useful processes for self-assessment by regulated parties (in this case, online platforms), but conducting an AIA can amount to a check-box exercise if they do not include more structured forms of accountability. **There should be a forum (eg. an independent body, a judicial forum or a regulator) to oversee the conduct of an AIA, as well as ensure the implementation of processes for mitigating algorithmic harms or risks that have been identified.**

References

Cohen, J. E., (2016). 'The regulatory state in the information age.' Theoretical Inquiries in Law 17(2) 369. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2714072

Gellert, R., (2015). 'Data protection: a risk regulation? Between the risk management of everything and the precautionary alternative', *International Data Privacy Law*, 5(1) 3.

Metcalf, J, 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

Seaver, N. (2019). 'Knowing Algorithms'. *digitalSTS: A Field Guide for Science & Technology Studies*, Princeton: Princeton University Press, pp. 412-422. Available at: https://doi.org/10.1515/9780691190600-028

Meaningful risk and impact assessments of algorithmic decision-making systems in the private and public sector

Andrew Strait, *Associate Director at the Ada Lovelace Institute*

How can regulators and developers of an algorithmic decision-making (ADM) system assess its risks and measure its impacts on individuals and society? More importantly, how can we trust these assessments are meaningful, rigorous, and transparent?

A variety of emerging algorithmic accountability mechanisms may provide the answer. Two of these mechanisms – algorithmic impact assessments (AIAs) and auditing – are promising methods for evaluating and documenting ADM system's risks and impacts throughout its lifecycle. The success of these methods in both the public and private sector will depend on the standardisation of evaluation methodologies, the creation of an underlying data access infrastructure, a successful integration with other regulatory and governance mechanisms, and the creation of soft and hard incentive structures for their use.

Drawing from the literature and Ada Lovelace's forthcoming work on methods for auditing and algorithmic impact assessments, this talk will address:

The importance of documentation: it's not exciting, it's essential! Impact assessments and audits produce artifacts that enable regulators and members of the public to apply scrutiny. Consistent, clear, and accessible documentation is crucial for tracing decision-making, clarifying an ADM's intended uses, disseminating information to downstream users or affected parties, and encouraging reflexivity from developers (Gebru, 2018; Mitchell, 2019).

Participatory approaches: impacts are an "evaluative construct" (Moss et. Al, 2021) that measure the effects of a technology, including what harms or benefits might arise. How impacts are defined will depend on who is involved in that process. A meaningful impact assessment must involve participatory engagement with affected stakeholders, including making these artifacts publicly accessible. (Ada Lovelace, 2021).

Incentives for use: private and public sector developers currently lack strong incentives to measure and evaluate impact. By focusing on particular "pinch points" in the ADM lifecycle like funding,

procurement, or market entry, regulators and policymakers can create forcing functions for different industry actors (e.g. engineers, researchers, product leads, legal counsel) to engage with these mechanisms.

Assessment throughout the lifecycle: ADM systems are dynamic and complex, raising novel challenges for their evaluation. Evaluations of a system in lab settings can fail to adequately address risks in context (Raji, 2021). Continuous evaluation of an ADM system at different stages of its lifecycle is essential to identify and mitigate potential harms (Raji, 2020). This will require novel regulatory powers and capacity, a marketplace of independent auditors, and clearly-defined requirements for different kinds of ADM systems.

Policy recommendations:

1. Require developers of ADM systems engage in continuous evaluation at all stages of the product lifecycle.
2. Require developers of ADM systems to engage in participatory methods for identifying stakeholders, defining harms, and evaluating impacts.
3. Mandate documentation standards and transparency reporting requirements.
4. Create clear reputational, legal, and performance incentives for engaging in accountability mechanisms.

Sources

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

Ada Lovelace Institute. (2021). Participatory data stewardship. Available at: https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/

Centre for Data Ethics and Innovation. (2021). User needs for AI assurance. UK Government. Available at: https://cdei.blog.gov.uk/2021/04/16/user-needs-for-ai-assurance/

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. arXiv preprint arXiv:1803.09010.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229).

Moss, E., Watkins, E.A, Singh, R., Elish, M.C, Metcalf, J. (2021) Assembling accountability: algorithmic impact assessment for the public interest. Available at: https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/

Raji, I. D., Denton, E., Bender, E., Hanna, A., Paullada, A. (2021). AI and the everything in the whole wide world benchmark. Neurips 2021, forthcoming. Available at: https://ml-retrospectives.github.io/neurips2020/camera_ready/18.pdf

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 33-44). Available at: https://arxiv.org/abs/2001.00973.