

The Government of Canada's Algorithmic Impact Assessment: Towards Safer and More Responsible AI

By: Noel Corriveau¹

Context: The Virtuous Equation

AI has quickly emerged as a disruptive technology that is rewiring the way we interact with the world. Government has to move at the speed of innovation and remain relevant, modern and responsive to the needs of Canadians. However, in the context of artificial intelligence it is critical for the Government of Canada (GC) to ensure a *virtuous equation*. On the one hand the GC needs to promote and encourage technological advances and we are making good strides in this direction.² But it is not all about innovation. Too often the discussions on AI focus only on innovation, efficiencies and the many promises of improved services and capabilities. What is often missing is the “ethics variable”. Ethics is the other core variable in our *virtuous equation*. Take for example the latest results from Ipsos Canada [Tech4Good survey](#):

- ONLY 22% of Canadians would find it acceptable to use algorithms in HR decision-making
- ONLY 28% of Canadians would find it acceptable to use AI to implement government policies such as who can immigrate into Canada
- ONLY 30% of Canadians find it acceptable to use algorithms to decide who can access government support such as employment insurance.

Ipsos Canada poll dated last May 2018

What this indicates is that the pace of technological advances is not aligned to social acceptability. You can have all the most advanced technology, but if trust is lacking, ultimately the value proposition fails. The Government needs to ensure that both ethics and innovation align.

Proposal: Algorithmic Impact Assessment

An impact assessment is a tool used for the analysis of possible consequences of an initiative with a view to provide recommendations as to how to deploy the initiative and

¹ Noel Corriveau, B.A., M.A., J.D. is counsel with the Federal Department of Justice on assignment at the Treasury Board of Canada Secretariat as a Special Advisor on Artificial Intelligence – Policy and Implementation. Noel.corriveau@tbs-sct.gc.ca

² For example the \$125 million Pan-Canadian Artificial Intelligence Strategy, the Supercluster initiative and the Innovation and Skill Plan are all examples of how Canada is cultivating advanced scientific research, talent and collaboration.

under what conditions. The merits of impact assessments predominantly are: 1) supports informed decision-making and the protection of societal concerns; 2) facilitates compliance with legal and other regulatory requirements (e.g. standards and directives); 3) when conducted in a transparent manner, appeals to public confidence and demonstrate corporate and social responsibility; and 4) being a “best-efforts obligation”, constitute evidence of due diligence which can potentially limit liabilities.

The Government of Canada is currently working on an Algorithmic Impact Assessment framework that would help institutions better understand and mitigate the risks associated with Artificial Intelligence by providing the appropriate governance, oversight and reporting, and audit requirements. In developing our proposed AIA Framework, we’ve underlined three key goals:

- Increase capacity to evaluate the impact of automated decision systems including legal and ethical issues, such a failure to monitor for unintended outcomes, or potential procedural fairness violations;
- Recommend the appropriate governance, oversight and/or design recommendations to institutions based on their existing or proposed automated decision systems; and
- Provide a mechanism for greater openness and transparency for public consultation and external review of the design and deployment of automated decision systems in the public sector, because we intend for the final results to be public.

How it works

The AIA is structured to be an electronic survey instead of the text-heavy approaches that have traditionally been used in the past. All results of the AIA will be exportable and available to the public. This will let researchers and civil society have a standardized means to monitor and compare which decisions are being automated and what are some of these systems’ core attributes. The AIA will be asking a wide variety of questions grouped into broad themes. For example:

- What is your system doing? Is it prioritizing files? Making decisions end-to-end?
- What types of decisions are resulting from this system? Is this a critical social program? A permit or licensing program? Food or product safety? Crime or fraud detection?
- Are the decisions based on objective criteria, or is there room for discretion to be showed? (Hint: We care about this a lot)
- Have you consulted everyone you need to (e.g. legal, HR, IT...)

- What are the processes you've put in place? How explainable are machine determinations? Are you monitoring for unintentional outcomes? Do you have a clear audit trail around decisions?
- Do you have a recourse mechanism in place in case people have questions or complaints?

Each answer is coded with five separate scoring categories that add or remove risk depending on the answers. The aim is that all of the scores are transparent, so you can see exactly where you accrue risk and where you can take measures to reduce it. We're looking to test five areas of interest:

- 1) The impact on individuals, businesses, and communities ("socioeconomic impact"),
- 2) The impact on government operations,
- 3) The complexity of the system,
- 4) Data management practices, and
- 5) Procedural fairness considerations.

Once you fill in your questionnaire, you'll be provided with an impact level from 1-4, which will determine which requirements will apply to your initiative. Requirements scale according to the risk level. I.e. peer review requirement for level 1 initiatives will be much less than those required for Level 4.

Our 3 Key Challenges: NIPS Proposal

There are three core areas that still require significant input. We would like to propose that each of these areas be considered for a workshop at NIPS.

1) Asking the Right Questions

We want to ensure that we are capturing all the right questions to properly assess the impact of algorithms. This includes data management questions, legal and ethical considerations, as well as social impact questions. To this end we believe NIPS would be a great opportunity to get feedback and explore new questions and areas of consideration.

2) Scoring the AIA

Scoring the AIA is currently our greatest challenge. The AIA needs an effective scoring methodology so that a proper risk level can be attributed. We have currently developed two scoring systems none of which are optimal. To this end, we believe presenting our current scoring methods to NIPS would be a great opportunity to revisit, co-create and experiment with alternative scoring methodologies.

3) Getting the Right Outcomes

It is critical that the AIA is successful in allocating the right risk level to AI initiatives. To this end, we are seeking to apply as many use-cases to the AIA as possible to test whether it is generating the right outcomes. To this end, we believe presenting the AIA in the context of NIPS would be a great opportunity for participants to test the AIA and run use-cases to see whether the outcomes align to expected results.

Thank you for your consideration.

Noel Corriveau, B.A., M.A., J.D
Special Advisor on Artificial Intelligence – Policy and Implementation
Treasury Board of Canada Secretariat
noel.corriveau@tbs-sct.gc.ca