



**RETURN BIDS TO:
RETOURNER LES SOUMISSIONS À:**

Bid Receiving - PWGSC / Réception des
soumissions - TPSGC
11 Laurier St. / 11, rue Laurier
Place du Portage, Phase III
Core 0B2 / Noyau 0B2
Gatineau
Quebec
K1A 0S5
Bid Fax: (819) 997-9776

**LETTER OF INTEREST
LETTRE D'INTÉRÊT**

Comments - Commentaires

Vendor/Firm Name and Address
Raison sociale et adresse du
fournisseur/de l'entrepreneur

Issuing Office - Bureau de distribution
Systems Software Procurement Division / Division des
achats des logiciels d'exploitation
Terrasses de la Chaudière
4th Floor, 10 Wellington Street
4th etage, 10, rue Wellington
Gatineau
Quebec
K1A 0S5

Title - Sujet RFI for AI Services	
Solicitation No. - N° de l'invitation 24062-190106/A	Date 2018-06-28
Client Reference No. - N° de référence du client 24062-190106	GETS Ref. No. - N° de réf. de SEAG PW-\$\$\$E-017-33657
File No. - N° de dossier 017ee.24062-190106	CCC No./N° CCC - FMS No./N° VME
Solicitation Closes - L'invitation prend fin at - à 02:00 PM on - le 2018-08-17	
Time Zone Fuseau horaire Eastern Daylight Saving Time EDT	
F.O.B. - F.A.B. Plant-Usine: <input type="checkbox"/> Destination: <input type="checkbox"/> Other-Autre: <input type="checkbox"/>	
Address Enquiries to: - Adresser toutes questions à: Lessard, Peter	Buyer Id - Id de l'acheteur 017ee
Telephone No. - N° de téléphone (613) 850-7602 ()	FAX No. - N° de FAX () -
Destination - of Goods, Services, and Construction: Destination - des biens, services et construction: Specified Herein Précisé dans les présentes	

Instructions: See Herein

Instructions: Voir aux présentes

Delivery Required - Livraison exigée See Herein	Delivery Offered - Livraison proposée
Vendor/Firm Name and Address Raison sociale et adresse du fournisseur/de l'entrepreneur	
Telephone No. - N° de téléphone Facsimile No. - N° de télécopieur	
Name and title of person authorized to sign on behalf of Vendor/Firm (type or print) Nom et titre de la personne autorisée à signer au nom du fournisseur/ de l'entrepreneur (taper ou écrire en caractères d'imprimerie)	
Signature	Date

Contents

1. Background.....2

2. Requirements2

2.1. Services2

2.2. Solutions3

2.3. Products3

3. Instructions To Respondents4

3.1. Response Costs4

3.2. Treatment Of Responses4

3.3. Enquiries5

4. Submission Of Responses5

5. Questions.....6

1. Background

The Treasury Board Secretariat of Canada (TBS) is seeking to understand how the Government of Canada and its Departments can leverage the benefits of artificial intelligence (AI)¹ to help improve service delivery and increase capabilities within their organizations.

The list of questions in section 3 below is meant to allow industry to:

- a. share views and inform the direction that the Government of Canada may take with respect to AI;
- b. comment on the risks associated with AI and mitigation strategies for reducing those risks; and
- c. inform the parameters that the Government of Canada may consider in determining scope and when or if an AI application is appropriate for achieving desired business objectives.

The information provided will aid in increasing the Government of Canada's understanding of the capability, viability and commercial availability of AI services, solutions and products. The industry perspective on the approach, business and general requirements, cost of development, viable business models and related contract and intellectual property considerations are of particular interest.

2. Requirements

TBS is seeking to develop a flexible procurement vehicle that would enable clients to move freely between services, solutions and products. The following list of categories, in conjunction with the documents provided with this Request for Information (RFI) represents the current understanding of the AI marketplace.

We would like to understand what the marketplace can offer in the following categories (services, solutions, products).

2.1. Services

a. Expert Advice on AI Applications and Taxonomies

Requirement:

- Provide presentations and overviews of the current state of AI technology, environmental scans and taxonomies of current AI applications; and
- Provide concrete used-cases of AI applications in the government context.

b. Feasibility Assessments

Requirement:

- Inform the parameters that departments should consider when determining scope for an AI application.
- Review work processes, procedures and decision flows and provide expert advice on whether there are opportunities to leverage artificial intelligence, machine learning and

¹ Recognizing that there is not a consistent agreement as to what constitutes artificial intelligence, including whether or not certain types of machine learning fall within the AI spectrum, for the purposes of this RFI respondents are asked to provide their responses considering a broad and inclusive approach to AI, including machine learning technologies.

predictive analytics. This could include providing recommendations, reviews and comments on the quality of existing data, proposed processes and applications.

- Comment on the risks associated with AI solutions and propose mitigation strategies for reducing those risks.

c. Peer Review

Requirement:

- Conduct quality control and peer review of data sets, algorithms and models.
- Assess and validate findings
- Review and comment on any project documentation
- Identify any risks and propose mitigation strategies.
- Publish and present finding, where appropriate
- Develop test scenarios and test scripts, as required.

2.2. Solutions

a. Design Implementation and Road-mapping

Requirement:

- Develop process maps and identify key phases for implementing an AI solutions
- Identify any ongoing requirements, such as necessary resource requirements and technical specifications.
- Provide expert advice and assistance in implementing the new processes and applications.
- Document workflows.
- Provide frameworks to measure results and efficiency gains.
- Participate in change impact analysis and change management activities. This includes developing awareness, training and communication.

b. Develop and Implement Pilots

Requirement:

- Develop pilots tailored to specific business requirements.
- This pilots could include applications that provide:
 - text analytics and sentiment analysis
 - clustering and pattern recognition
 - outcome predictions
 - chatbot interactions
 - automated decision making support

2.3. Products

Requirement:

- Describe current commercially off-the-shelf products (COTS) related to AI. The applications can relate to:
 - text analytics and sentiment analysis
 - clustering and pattern recognition
 - outcome predictions
 - chatbot interactions
 - automated decision making support

3. Instructions to Respondents

Nature of Request for information

This is not a bid solicitation. This RFI will not result in the award of any contract. As a result, potential respondents of any goods or services described in this RFI should not reserve stock or facilities, nor allocate resources as a result of any information contained in this RFI, nor will this RFI result in the creation of any source list. Therefore, whether or not any potential respondent responds to this RFI will not preclude that respondent from participating in any future procurement. Also, the procurement of any of the goods or services described in this RFI will not necessarily follow this RFI. This RFI is simply intended to solicit feedback from industry with respect to the matters described in this RFI.

a. Format of Responses Requested

- i. Format: Respondents are requested to submit one soft copy of their response in Portable Document Format (PDF).
- ii. Cover Page: If the response includes multiple volumes, respondents are requested to indicate on the cover page of each volume the title of the response, the solicitation number, the volume number and the full legal name of the respondent.
- iii. Title Page: The first page of each volume of the response, after the cover page, should be the title page, which should contain the:
 - a) title of the respondent's response and the volume number;
 - b) name and address of the respondent;
 - c) name, address and telephone number of the respondent's contact;
 - d) date; and
 - e) RFI number.

b. Part A should include feedback to the Requirements depicted in Section 2.

c. Part B should include answers and feedback to the questions listed in Section 5.

d. Numbering System: Each question has its own unique number. Respondents are requested to prepare their response using the numbering system corresponding to the one in this RFI, and to repeat the question prior to their response for ease of reviewer reference. All references to descriptive material, technical manuals and any brochures included as part of the response should be clear both in the citation and on the referenced document. All should be referenced accordingly.

3.1. Response Costs

The Government of Canada will not reimburse any respondent for expenses incurred in responding to this RFI.

3.2. Treatment of Responses

a. Use of Responses: Responses will not be formally evaluated. However, the responses received may be used by the Government of Canada to develop or modify procurement strategies or any draft documents contained in this RFI. The Government of Canada will

review all responses received by the RFI closing date. The Government of Canada may, in its discretion, review responses received after the RFI closing date.

b. Review Team

A review team composed of representatives of TBS will review the responses. The Government of Canada reserves the right to hire any independent consultant, or use any government resources that it considers necessary to review any response. Not all members of the review team will necessarily review all responses.

c. Confidentiality

Respondents should mark any portions of their response that they consider proprietary or confidential. The Government of Canada will handle the responses in accordance with the Access to Information Act.

d. Pre-Submission Industry Day Session

The Government of Canada will host an industry day session on July 20, 2018, in Ottawa, for the purpose of explaining its requirements and to allow industry to ask questions and seek clarifications. Industry Day information will be posted on Buy and Sell by July 6th, 2018.

e. Post-Submission Review Meetings

The Government of Canada may, in its discretion, hold a single Post-Submission Review Meeting with all interested vendors or request individual Post-Submission Review Meetings with selected respondents to provide clarity on information provided, or to invite a presentation about some or all of the proposed solutions. If required, these will be held at the most appropriate location, to be determined at a later date. The intent of these meetings will be to provide an opportunity for a face-to-face discussion with respondents. Although respondents may request a meeting, and their request will be considered, the Government of Canada will determine whether or not it requires additional information from any given respondent and will schedule meetings accordingly. All such requests, by respondents, should be forwarded to the Contracting Authority identified herein. Note that a maximum of two (2) hours will be set aside for any meetings with respondents.

3.3. Enquiries

Because this is not a bid solicitation, the Government of Canada will not necessarily respond to enquiries in writing or by circulating answers to all potential respondents. However, respondents with questions regarding this RFI may direct their enquiries to the Contracting Authority identified herein.

4. Submission of Responses

- a. Respondents should send responses electronically via e-mail to the Contracting Authority's e-mail address identified herein by the date specified on the front page of the RFI.
- b. All requested information is to be provided to the Contracting Authority on or before the closing date of the RFI.

Contracting Authority
The Contracting Authority for this RFI is:

Peter Lessard
E-mail Address: peter.lessard@tpsgc-pwgsc.gc.ca
Telephone: 613-850-7602

5. Questions

This document has a number of attachments:

- Annex A: Responsible Artificial Intelligence in the Government of Canada
- Annex B: Algorithmic Impact Assessment (AIA)
- Annex C: To be provided with the July 6, 2018 RFI amendment

Please take into account the contents of all documents when considering the following questions.

a. AI Market Overview

1. Given the information provided to you are there any significant gaps in the identified requirements or background information? If so, how could they be improved?
2. Can you please describe how the AI industry is currently structured?
3. What are some of the key considerations in relation to how AI services, solutions and products are currently being provided by the industry?
4. What is the best way to successfully engage and work with the industry?

b. Company Overview

1. Please provide detailed information about the services, solutions and/or products provided by your company.
2. Please describe costing models for your services, solutions and/or products.
3. Has your company provided AI/ML services, solutions or products for public sector or private sector organizations in Canada or elsewhere? If so, please describe.

c. Intellectual Property

1. Given the importance of transparency, what are some of the key considerations that pertain to licenses and IP in the context of AI services, solutions and products?



Responsible Artificial Intelligence in the Government of Canada

Digital Disruption White Paper Series

Version 2.0
2018-04-10

Responsible Artificial Intelligence in the Government of Canada

1. Table of Contents

1. Table of Contents	1
2. Version History	2
3. Executive Summary	3
4. Introduction	4
4.1. Objective of this paper and intended audience	6
4.2. Automation and Artificial Intelligence	7
4.3. Narrow and General Intelligence	10
5. AI for Smarter Government	10
5.1. AI for the Delivery of Services to the Public	10
5.1.1. Smarter Search	11
5.1.2. Chatbots	12
5.1.2.1. User Experience Considerations	14
5.1.3. Automated Decision Support	15
5.1.3.1. Appropriateness of Automation	16
5.1.3.2. Transparency and Recourse	17
5.2. AI to help design policy and respond to risk	18
5.3. Applying AI to the internal services of government	19
5.3.1. Information Management	19
5.3.2. Automated Content Generation	19
5.3.3. People Management	20
5.3.4. Security and Access Management	20
6. Policy, Ethical, and Legal Considerations of AI	21
6.1. Ensuring High-Quality Data	22
6.1.1. Prevention of Data Bias	22
6.1.2. Data for Insights and Privacy Rights	24
6.2. Transparency and Accountability	25
6.2.1. Accounting for the Actions of AI: The “Black box” Problem	25
6.2.2. Model Design and Outcome Biases	Error! Bookmark not defined.
6.2.3. Social Acceptability	26
6.3. AI and the Law: An Emerging Landscape	26
6.4. Technical Considerations	27
6.4.1. Cybersecurity considerations	27
7. Rethinking a Post-AI Enterprise	28
7.1. New Approaches to the Workforce	28
7.2. Evolving How Government Works	29
8. Conclusion	30

Responsible Artificial Intelligence in the Government of Canada

2. Version History

#	Date	History
Sections 0.1 – 0.10 –Early Concept		
0.1	July 21	Release
0.11	August 7	Release
0.12	August 17	Release
0.2	September 22	Release
0.3	October 2	Release
0.4	October 16	Release
Sections 1.0 – 1.X – Working Drafts for Broad Consultation		
1.0	October 27	First draft for open comment
1.1	November 6	New introduction, new section on security and user access control
1.2	November 21	New sections on inclusion and cybersecurity, revisions to sections on AI for policy, revisions to “evolving how government works,” new box on anthropomorphism, risk test appendix removed - it belongs more in directive format than in this white paper.
1.3	December 7	Version shared with FNIGC, several companies
1.4	March 5	Version sent to the Privacy Commissioner
Sections 2.0 - Senior Management Consideration		
2.0	April 17	Version for formal translation

Responsible Artificial Intelligence in the Government of Canada

3. Message from the Chief Information Officer of the Government of Canada

(to come)

4. Executive Summary

Artificial intelligence (AI) is a term used to describe a suite of related technologies intended to simulate and enhance human cognitive capabilities, such as pattern recognition, judgement, vision, or hearing. Having first been conceived in the 1940's, AI has advanced rapidly in recent years due to a combination of vast quantities of data, new mathematical techniques, and inexpensive computing power. AI systems now underpin many of the consumer products that Canadian use on a daily basis, from curating what media content we consume based on our interests, to helping us navigate our towns and cities. There are real-world examples of AI systems operating vehicles, writing newspaper articles, or generating art, that challenge previous assumptions of the types of tasks that can be delegated to machines.

Just as AI systems are rapidly transforming the world around us, so too is it expected that AI will transform the way that government operates. Imagine virtual service agents assisting Canadians and businesses with completing routine transactions 24 hours a day, seven days a week. AI systems can monitor the status of industries to detect early warning of regulatory non-compliance. They can sift through, structure, and recombine vast stores of data to help government institutions understand the information that they currently have, in order to more intelligently design public policy. These technologies have the potential to guide the public service towards a future of greater effectiveness and responsiveness to the needs of society than was ever possible before.

While the power that AI systems may bring to government could be significant, they must be deployed in a responsible and ethical manner. AI systems often require “training” using datasets that are reflective of the problem needing to be solved. If these data were collected or tabulated in a way that carries bias, then the outcome will be AI recommendations or decisions that are biased as well. Further, some AI systems currently operate as “black boxes,” meaning that the decisions they make are difficult to audit or fully comprehend. In light of these limitations, it is important to understand where it is appropriate to deploy different types of AI systems, balancing the potential for gains in efficiency and effectiveness of government with the risk of misuse. Finally, although AI will afford institutions with new capabilities, institutions will need to apply a strong ethical lens to whether the technology should be deployed at all in certain circumstances.

Responsible Artificial Intelligence in the Government of Canada

AI is a capability that rests atop an expert and disciplined data science practice within institutions, as well as leveraging Canada's leading AI talent base. These systems will challenge how government institutions work, demanding a prioritization of good data governance practices, and requiring new skillsets of knowledge workers.

This paper proposes a set of seven principles that will be expressed in all future Treasury Board policy on the use of AI systems in government:

1. People should always be governed – and perceive to be governed – by people;
2. AI systems deployed on behalf of government should be trained to reflect the *Values and Ethics of the Public Sector* as well as Canadian and international human rights obligations; they should be used to reinforce these values where possible;
3. Organizations are accountable for the actions of AI systems, and should build systems that are auditable;
4. Understanding the need to protect privacy and national security, AI systems should be deployed in the most transparent manner possible;
5. Organizations should ensure that reliable contingencies are in place for when AI systems fail, or to provide services to those unable to access these systems;
6. AI systems should be developed in a diverse team that includes individuals capable of assessing the ethical and socioeconomic implications of the system;
7. AI systems should be deployed in a manner that minimizes negative impact to employees where possible, and should, where feasible, be created alongside the employees that will work with them.

5. Introduction

First it was chess, then Go, then poker. One by one, we have taught machines to exceed us in some of our most treasured – and complicated – games. These accomplishments showcased advancements in techniques achieved much faster than predicted, and were at least partially responsible for kicking off an era of massive investments and excitement in artificial intelligence. We have trained machines to mimic the outcomes of human learning and decision processes, such as adaptation, bargaining, and bluffing. With successive and public displays of computing prowess by the likes of IBM, Deepmind, or Facebook, and the rapid growth of a startup ecosystem, advances in AI have begun to dominate the press and capture the public's imagination.

Responsible Artificial Intelligence in the Government of Canada

While AI was originally conceived in the 1940's, over the past decade, these applications have been deployed in such variable and extensive ways that it increasingly drives the modern economy. AI has replaced humans on stock market floors¹ and in the management of multi-billion dollar hedge funds.² It assists with medical diagnoses and operates complex machinery autonomously. It has been applied to corporate process and workflow automation to increase efficiency of their operations. AI agents are beginning to use natural language effectively enough to interact with humans via intelligent chatbots. There is a very high likelihood that by 2025, AI will touch every aspect of modern society in ways both visible and invisible to Canadians.³

Since the 1970s, early investments in Canadian researchers allowed an AI industry to bloom here. The advances of Canadian pioneers in machine learning positioned this country as a global leader in AI research, development, and application. Budget 2017 committed \$125 million to launch a Pan-Canadian Artificial Intelligence Strategy to support these clusters and attract the talent they need to maintain their advantage. Establishment of superclusters in Montreal, Toronto, and Edmonton has seen both the rise of world-leading research institutes as well as an ecosystem of AI startups that are internationally competitive and driving innovation.

Now, the Government of Canada is looking into how it can harness the opportunities provided by AI to offer novel and more timely services to citizens and other users,⁴ as well as improve the effectiveness and efficiency of its operations. Federal institutions are working towards offering better user experiences to make their services easier to use, but these gains will not accomplish a frictionless service environment if the person faces weeks-long backlogs in having a benefit application processed. Especially in circumstances where work is routine, AI systems can work faster and often more consistently than humans performing the equivalent tasks, and will work over evenings, weekends, and statutory holidays. Their capacities for decision-making are not adversely affected by physical fatigue or the natural emotional and relational situations people face based on their natural makeup. AI systems can be deployed by service institutions to answer questions posed by users – as well as make eligibility determinations – in order to dramatically improve the response time of service.

On the other hand, when administrative tasks are complex and value-laden, it can be difficult to ensure that the actions of the AI systems align with the spirit and intentions of the policy being implemented. Working with complex social and economic systems is considerably more complex than a game of Go. How do we know whether an AI system is appropriately trained for its task, and that data is interpreted in a manner that is accurate and responsible? How do we know whether AI is making biased or prejudicial

¹ See example: <http://www.bbc.com/news/business-34264380>

² See example: <https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-ai-artificial-intelligence-management>

³ A qualitative survey by the Pew Research Center of over 2,500 academics, policy analysts and corporate executives found broad consensus to support this prediction. While the study was American, respondents were international. See: Pew Research Center, “AI, Robotics and the Future of Jobs.” Link: <http://www.pewinternet.org/files/2014/08/Future-of-AI-Robotics-and-Jobs.pdf>

⁴ This paper uses the term “users” to represent the diverse groups that use Government of Canada services including, but not limited to, citizens, permanent and temporary residents, and businesses. It avoids the term “client” to reduce confusion with the legal term.

Responsible Artificial Intelligence in the Government of Canada

decisions? How can AI systems be coded to meet similar legal obligations as human public servants, such as the Charter of Rights and Freedoms or the *Privacy Act*, and who is responsible when they fail to meet these obligations? How do we teach it social, cultural, or geographical context such that it can make decisions in a nuanced fashion? How do we know the rationale behind the decisions of an AI system? What types of decisions should always require some form of human intervention? How do we know that the data on which an AI system is trained, which is sampled from real data about real Canadians, is kept secure and private once the AI system is in deployment? What are the workforce requirements in a post-AI world?

Governments worldwide are now grappling with the consequences of a technological development that is transforming service delivery across sectors. The United States, United Kingdom, France, the United Arab Emirates, China and Japan are just some of the jurisdictions that have undertaken high-level examinations of AI systems within their respective governments and on their economies writ-large. The Government of Canada has the opportunity to build on the brain trust of private sector and academic leaders in this field to position itself as a world leader in AI for policy development and service delivery. It has the opportunity to signal to all sectors that AI can be harnessed in a manner that is ethical and supportive of positive outcomes for Canadians without sacrificing the benefits of the technology.

While AI is undergoing rapid advancement, it is important that the policy, ethical and legal implications of the use of this technology to deliver government services be addressed methodically and with an understanding of this complexity. The service delivery opportunities are significant, as are the pitfalls.

5.1. Objective of this paper and intended audience

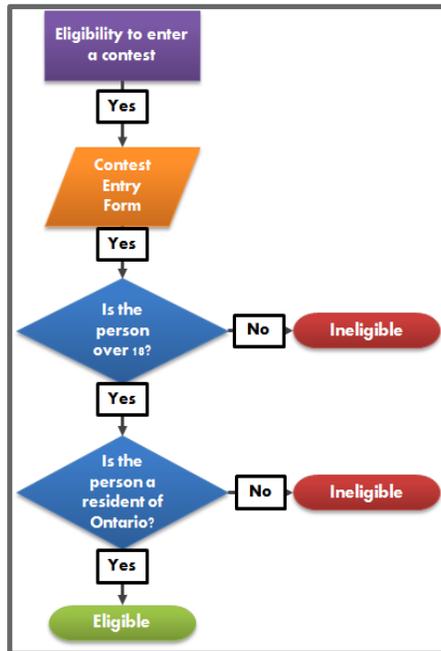
The scope of this paper is limited to the specific use of AI applications by federal institutions for their own use only; it does not touch on the Government's response to automation in the private sector and its effect on society. This scope is broadly aligned with the mandate of the Treasury Board in its role in setting general administrative policy for federal institutions.

This white paper will examine the policy, ethical, technical, and legal considerations around the use of this technology within the Government of Canada. Its primary objective is to assist federal institutions by providing recommendations on how these systems should be implemented. The intended audience is therefore broad, from Deputy Heads or Chief Information Officers wishing to understand a significant new technology, to policy managers or service designers looking to apply AI to the programs or services that they provide. At the same time, it is intended to communicate to the AI development ecosystem in the academic and private sectors the use cases and policy considerations that are common in the federal government.

Throughout the paper, illustrative examples are used to show how this technology can be beneficial to users. Unless otherwise specified, these examples do not represent any existing plans of the Government of Canada and should be considered theoretical only.

Responsible Artificial Intelligence in the Government of Canada

5.2. Automation and Artificial Intelligence



Humans have always been intrepid designers of tools. From the scythe and wheel to the internal combustion engine and the computer, we have always designed tools to produce more from less. For most of human history this has led to technologies that have extended our physical capacities, but with the outbreak of the Second World War, humanity started designing tools that started to extend our cognitive and analytical capacities as well, such as memory, attention, judgement and decision-making. In a sense, we started designing brains for our tools.

We eventually designed tools that took over tasks for us completely. Automation has been a hallmark of industrialization since the robot Unimate was deployed in a New Jersey GM plant in 1961 for hazardous die casting, not just for physical tasks, but for analytical ones as well.

Behind the automated processes that drive the 21st century economy are a series of logical instructions known as algorithms.

Like a recipe, algorithms are processes that inform a machine how to perform a specific task. They can often be broken down into a series of decisions that are defined by the programmer; such as “is the individual over 18 years old?” or “is the individual a legal resident of Ontario?” The output is decided based on these decisions. The rules of these algorithms do not change unless programmers decide to change them. Closed-rule algorithms are used in the support of decisions widely in the private and public sectors today; for example, the Canada Revenue Agency uses closed-rule algorithms to support tax processing, with the rules defined by legislation and regulation.

Enter Artificial Intelligence

While it was the eminent British computer scientist Alan Turing that first conceived of “the thinking machine,” the term “artificial intelligence” was coined later in 1956 by the American computer scientist John McCarthy to describe “the science and engineering of making intelligent machines.” As technology has evolved, AI has grown to become a term that includes a broad spectrum of related technologies that seek to imitate and enhance aspects of human intelligence, such as vision, identifying patterns in information, or understanding language. In a sense, AI is when computers do what only humans could do before. The term is used to describe applications as innocuous as a system that recommends books to read, to fictional advanced human-like intelligence capable of everything a human is. As such, there is no single, internationally-recognized definition for AI, and the term may mean different things to different people.

The development of **machine learning** was a critical milestone. Machine learning is a method by which algorithms can be trained how to recognize patterns within information, and the ways in which data interrelate. For example, a learning algorithm that recommends books based on your purchasing history

Responsible Artificial Intelligence in the Government of Canada

provides better recommendations as you purchase more books. It does this without a human on the backend needing to adjust the programming instructions. If that algorithm had access to your browsing history as input data - and assuming that it was programmed to know what to do with that data - its recommendations may improve even more because it begins to “know” your tastes better.

Machine learning is by no means the only application of artificial intelligence. **Natural language processing** allows computers to parse meaning and context out of written text. This is used extensively, for example, in legal analysis software to derive insights from large volumes of text. **Machine vision and hearing** provide machines with the capability of structuring, and using, typically unstructured data such as imagery or sound. This is used in a diverse range applications, from autonomous cars “seeing” obstacles to smartphone applications that can identify a song played in public.

Either one or a combination of these techniques underpin many of the private sector digital services that people use regularly worldwide. Major social networking platforms, media platforms, and smartphones all run machine learning algorithms that provide services such as navigating traffic or curating news. It is not necessary to use machine learning in all approaches to automation; for applications where rules are precisely defined (such as the example above), a closed-rule algorithm is sufficient for the task.

Early experiments have existed since the late-1950s to show how machines are capable of learning and self-improvement. Today, researchers and developers have access to powerful and inexpensive cloud computing resources, parallel computing, as well as profoundly more data. Smartphones and the sensors located within them, coupled with the popularity of social media and internet culture means that a typical person produces a bounty of harvestable data every day - even when they are sleeping.⁵ As a result, the development - and implementation - of AI has progressed rapidly in the last ten years. As the Internet of Things connects common consumer products and appliances to the internet, the data points that we generate in our day-to-day lives will likely grow exponentially.

This ability to capture and use data in unprecedented ways has had a direct impact on the development of AI because of these technologies’ need for sufficient quality and quantity of data. Think of AI as a very sophisticated engine; without data to fuel it, it can’t propel the vehicle. Data needs to be available in sufficient quantity, they need to be relevant enough to the task at hand, they need to have been collected and described in a manner that is free of bias, and they need to be in a format that is readable by a machine. Despite addressing AI, much of this paper is devoted to issues surrounding data rather than the instructions precisely because insufficient quality and quantity of data can render the most expertly-programmed AI useless - or worse - harmful.

We are now at a point where machine learning can enable AI not only to replicate many human tasks - it can come close to surpassing our effectiveness at certain tasks, such as recognizing subjects of images,⁶ or reading lips.⁷

⁵ For example, by using an app that monitors sleep time and quality.

<http://ns.umich.edu/new/multimedia/videos/23822-smartphones-uncover-how-the-world-sleeps>

⁶ Based on 2017 results of the University of Washington MegaFace challenge:

<http://megaface.cs.washington.edu/results/facescrub.html>

⁷ Based on LipNet results. See: <https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/>

Responsible Artificial Intelligence in the Government of Canada

Advances in techniques

There are many approaches that developers take to AI; for example, **deep learning**, a branch of machine learning, has been used extensively in modern private sector services. While many deep learning algorithms use labelled data, it also brought the capability of using unstructured data such as audio or visual data, allowing the system to extract features of information on its own.

There have been significant advances in **artificial neural networks** in recent years. Inspired by the human brain, neural networks are composed of artificial neurons, which receive data individually and calculate outputs independently, allowing a complex problem to be broken down into millions of simple problems and then reassembled as one answer. As the network is provided more data, it can identify new and complex relationships in data, much like how the human brain forms synapses. This complex relationship is encoded in the weights, learned during model training, that connect the neurons in the neural network.

For example, rather than just learning what a bear is based on analyzing millions of images tagged as “bear,” a deep learning AI can extract features from images of a bear on its own. Humans do that as well; we learn a bear’s size and shape, where a bear may be found, typical colours of its fur and its family structure. That way, when we see an image of a bear that we have never seen before, we can infer that what we are seeing is a bear based on understanding its components.

The complication of deep learning is that it is not always possible to have access to massive data and to understand the importance associated with different variables of the problem. Using the above example, it is very difficult to understand whether an AI neural network considers fins as important than scales in determining whether something is a bear or not, both because the network is complex, but also because as the network is exposed to more examples of bears, this weighting may change. This process is often reliant on very large volumes of data that are broadly representative of the world within which the system will operate; for example, an autonomous vehicle trained exclusively in the UK could not be deployed in Canada, where driving is on the opposite side and some rules differ.

Another approach is, reinforcement learning; this is a subset of machine learning whereby machines are trained by being rewarded for desired outcomes and punished for undesired ones, similar to how we train dogs to play fetch. Rules are provided to the algorithm as to what it must do to earn a reward; for example, if the bear brings the ball back, it will get a fish to eat. The bear will not receive a fish if it does not bring the ball back. Reinforcement learning is especially useful in situations with well-defined outcomes, for example games and puzzles.

Reinforcement learning algorithms can be trained in advance using simulations, but they adapt more quickly once able to interact with its intended operating environment. However, they need clear definitions of “right and wrong” - outcomes that are desirable or undesirable, and the choices of those definitions are laden with values.

The choice of methodology will matter depending on the problem needing to be solved.

Responsible Artificial Intelligence in the Government of Canada

5.3. Narrow and General Intelligence

Whereas you are a multifaceted individual with a number of potentially unrelated interests, AI is often targeted for a single objective or task. This is known as “narrow” intelligence; while it can excel at one task – even surpassing a human – it cannot learn a second task without being explicitly targeted to do so. For example, while you may be a software engineer that speaks four languages fluently and is an amateur chef, an AI system trained to identify high-risk travellers cannot simply choose to learn to translate languages. This is because AI is software and does not have agency.

While research is underway to determine whether AI can achieve general intelligence, this achievement is still highly theoretical. A generally intelligent AI brings with it significant policy implications as well, but this paper will focus on the implications of narrow AI.

AI is software, not an organism

For decades, science fiction has introduced AI characters – whether in robot or incorporeal form – to the social consciousness. The popularity of characters like HAL 9000 or C-3PO may cause us to ascribe some degree of personification to AI. While it is designed to mimic human intelligence, the “learning” and “understanding” that a machine undergoes is different than the biological processes that we humans rely upon.

This paper refers to AI using humanlike semantics from time to time because it is a helpful way to communicate technical concepts, but it is important to remember that fundamentally, AI is software, not a conscious being, and should not be ascribed agency over its actions. Doing so removes the accountability of an organization over its software.

6. AI for Smarter Government

AI is not a technology looking for a problem; it is a suite of tools with the potential to help the GC deliver services more effectively, design policy more responsively, and potentially enable an entire suite of new capabilities in designing policy and delivering services. As the set of applications is diverse, its potential impact on the public sector is wide-ranging. Institutions have been examining applications that can be organized into three interdependent themes:

1. Applying AI to the delivery of services to the public
2. Applying AI to help design policy and respond to risk
3. Applying AI to the internal services of government

6.1. AI for the Delivery of Services to the Public

End-to-end digital self-service is the norm throughout much of the private sector service spectrum. The ability to access the entire continuum of the service from application to delivery without the need for a paper form, or for the user to have to interact with a service agent, is typical. Ideally, the service experience from authentication to application to receipt of benefit or issuance of payment should be a

Responsible Artificial Intelligence in the Government of Canada

seamless process that does not require the use of a phone or visit to a service centre unless chosen as the preferred way to receive service.

The government has decided to prioritize the development of digital services. Phone and in-person channels are inherently less convenient for users, as opening times are restricted, require waiting on hold or in line, or involve travel times. For individuals and businesses alike, lengthy wait times, or the requirement to access services during business hours can lead to an unacceptable loss of leisure time or productivity. Assuming that the digital service offering is understandable, convenient, and accessible enough for someone to want to use it, there is incentive for all parties of a service transaction to want to move to the digital channel. According to the Canadian Radio-Telecommunications Commission, broadband access in Canada will likely reach 90% as soon as 2021, and digital services will be more within reach for the vast majority of Canadians.⁸

Even if all services were provided digitally, as of today, there may be services by which some will elect to use other channels. There are some complex or sensitive needs that may demand more nuanced or personal service provision. Some people simply may feel more comfortable raising their issues in front of another person. In these circumstances, people may use an alternative channel such as phone or in-person if these are accessible to the individual. Even in these cases, AI can empower services by providing faster decisions, or tools that provide an overview of the individual's sentiment during the progress of the call.

More intelligent digital tools interacting directly with a user can play a role in keeping them on the digital channel. Smarter search and chatbots are capable of parsing natural language into searchable terms, accessing information located in FAQs, manuals or even specifically-identified internal documents and reply to the question in a way the user can understand. With additional information and user feedback, these tools will continuously improve at this task without the need for direct human intervention.

6.1.1. Smarter Search

Building a website targeted at millions of people presents a challenge; people interpret information differently, and may have different expectations as to where information can be found. Usability testing can help understand how people are interpreting information on a website, but advances in natural language processing (NLP) make the task of finding relevant information much easier than it used to be.

NLP technology parses natural language into underlying meaning, which then can be used in service of some task. For example, if a user loses their job, rather than having to look up Employment Insurance specifically, search for "I've lost my job" and see results that are relevant to that request. Over time, the application learns the relevance between search statements and the services that people are looking for. This is superior to older search methodologies, which would literally scan for the statement "I've lost my job" in web content. Over time, the algorithm will learn more patterns and do a better job at understanding what users want. NLP search functionality is widely used in the private sector today.

⁸ CRTC <http://www.crtc.gc.ca/eng/internet/internet.htm>

Responsible Artificial Intelligence in the Government of Canada

6.1.2. Chatbots

Chatbots are virtual user service representatives that offer capabilities of searching for information, or escorting a user to the right webpage. They work similarly to NLP search, but add a layer of interactivity and personalization.

The capabilities of an AI chatbot can be scaled up over time to provide expansive levels of user care as it gains experience and improves the way it manages information. It can offer responsive services, answering queries related to services passively. Eventually it can expand to become more navigational, offering hints, advice, or step-by-step instructions more reflexive of where a person is in the continuum of their service experience. Eventually an AI can be capable of actually executing instructions, such as accessing and pre-filling a form based on natural language. However unlike a website where a “what’s new” section can easily communicate new information or services available, some thought must be given to how the end user is aware of new functionality of a chatbot.

A chatbot may be offered to clients embedded in your webpage, or within another platform where your users are commonly found, such as SMS text messaging, Facebook Messenger, WhatsApp, Twitter, or Slack. This technology has advanced significantly over the past five years and is expected to continue its rapid advancement for the next decade, for providing both external and internal services.

Chatbots offer a diverse opportunity to provide services to users. Chatbots help filter routine questions away from human service agents so that they may focus on helping users through complex or distressing cases, or cases where a user is uncomfortable relaying their circumstances to a machine. They may also assist with public consultations on policies or programs, by being able to ask follow-up questions and react to user feedback in a much more nimble fashion than a survey.

This technology has been deployed successfully in the public and private sector. The United States Citizenship and Immigration Services uses a chatbot named Emma to answer users’ questions and provide a pre-check for eligibility. Emma not only answers questions, but provides navigational services; the search query “I’ve been offered a job in the US” not only provokes a response from Emma, but brings the user to the “Working in the United States” site. The bot is trained in English and Spanish. Another bot, Sgt. Star, is deployed by the US Army to answer questions to prospective recruits.

Institutions looking to deploy chatbots will need to ensure that there is training data available for the bot to learn the appropriate terminology for the service. This data can include previous interactions with clients looking for the service in question, whether emails, chat logs, transcripts from phone conversations, or social media. Ideally the datasets would include data on the outcome of the service interaction as well to ensure that responses to questions are those that actually satisfy clients.

Chatbots have limitations. As described above, conversations carry a lot of information outside of the basic text. Emotional queues or the use of sarcasm and humour can quickly confuse an AI conversational agent, or teach it bad behaviour. While they are adept at managing basic questions, a lengthy, interactive conversation is not possible at this time. Some chatbots provide a user with a defined set of potential inputs to reduce errors in the conversation, which results in a more scripted interaction. This can be useful

Responsible Artificial Intelligence in the Government of Canada

for quickly helping users find the information they need, although scripted interactions quickly become difficult to control as the scope of the bot's responsibilities increase.

An additional benefit of chatbots is their ability to structure data through a standardized approach to collection. Through interactions with users, a chatbot can help reduce spelling errors, inappropriate entry of dates and addresses, etc. This improves overall data quality, which in turn could help eligibility determination.

Chatbots offer transactional capability as well, merging the functions of both a virtual front-line service agent and the application form by collecting information directly from the user or their file in the institution's Customer Relation Management software.

It's important to remember that a user interacting with a chatbot may ask questions that are well outside the scope of its expertise. Users may disclose important personal information even when advised not to; they may even require immediate emergency assistance. In such a circumstance a human would be guided by a mix of their training and their own moral compass, but machine intelligences would need a means to triage these events, as well as pre-programmed responses.

Just like a human agent, a chatbot needs to be treated as an agent of the organization, which means that the information that it provides must always be accurate and up-to-date. Learning chatbots may provide advice to Canadians and, like humans, sometimes make mistakes. For example, a chatbot may give a person the wrong form or provide them an incorrect deadline. Chatbots that are designed to actually replace a form through conversational means may misinterpret input and submit incorrect information.

There have been significant and swift advances in chatbot technology, but despite these advances, it is a long way from flawless. In the future, bots have the potential for replacing forms as a way to collect information from users. They may even emerge to become the primary service delivery platform. Assuming that they have access to the widest range of information possible, bots can theoretically inform a user about any service in any institution with an almost expert like knowledge, far surpassing the ability of one individual's recall.

Finally, there are those in Canada who do not have access to reliable broadband internet, and may not in the near future. It is important that institutions continue to cater to these users and do not solely rely on chatbots for front-line services.

Is your institution ready for a chatbot?

When determining whether to deploy a chatbot, an institution should be able to answer the following questions:

Is there a clear business driver for the chatbot? Does your institution receive a high volume of routine inquiries?

Are the most common inquiries known and are data available to answer them?

Responsible Artificial Intelligence in the Government of Canada

What can be automated without taking away from the user experience and satisfaction?

What is the sensitivity of the information that the chatbot will likely receive or relay?

Will the interaction be an entirely scripted one, or allow the user to ask open questions?

Will there be an escalation process to a human live chat?

Does your institution have staff ready and able to provide ongoing training and direction to the chatbot?

Can interactions be stored in your CRM? Will it enable engagement across other channels (e.g. email, phone, in-person)?

6.1.2.1. User Experience Considerations

The GC has a wide policy and service landscape; if chatbots speaking to these policies and services offer interaction experiences that differ significantly, then users' acceptance of this technology can suffer and benefits will be unrealized.

A chatbot should not be used as a substitute for good discoverability of information on a website; it can add supplementary information or clarification to a user, but should not be seen as to replace the need for a well-designed site.

Chatbot conversations should be introduced with a brief privacy notice that is compliant with the *Treasury Board Standard on Privacy and Web Analytics*. This notice should provide a link to a page with more information on the information collected in the course of the conversation, including any metadata, for example: time and date, duration, whether the conversation was ended by the user or the agent, whether and when the discussion was escalated to a human, etc. Additionally, users should be informed that they are communicating with a chatbot.

Bots should be able to relay information in a professional tone as a representative of the Government of Canada. Machine learning chatbots may learn language that is potentially unprofessional, abusive, or harassing if exposed to sufficient examples. Where possible, institutions should work with vendors to prevent them from learning this behaviour, whether using a keyword blacklist, or other methodology. It is important to be continually monitoring chatbots' performance in this regard.

Responsible Artificial Intelligence in the Government of Canada

Some institutions may choose to use avatar, which is a personification of the chatbot. Visual avatars that express some emotional range improve users' belief in the competence of the virtual agent.⁹ The question of whether or not a chatbot should be gendered as male or female - or, for that matter, anthropomorphized (meaning: made to appear human - deserves close attention. It is unclear whether the use of a female gendered "assistant," could serve to perpetuate false, misleading and ultimately harmful cultural stereotypes about the status of women. To avoid a misstep in this sensitive area, some organizations have made the proactive decision to characterize their assistants as androgynous, such as Capital One's *Eno* and Sage's *Pegg* or non-human, such as Google's *Voice Assistant*.¹⁰

Institutions should be mindful that people in rural or remote locations may encounter latency that will affect their ability to respond to the chatbot's queries. It's important to ensure that response times from the user are permissive.

Chatbots must be accessible and meet accessibility standards and requirements of the GC. It is also important that chatbots be able to be read by screen readers, or are able themselves to communicate vocally, for persons with visual disabilities.

They should use plain language so as to be understood by users with varying levels of education or comfort with Canada's official languages. There is an opportunity to offer chatbots in a wide variety of languages should enough training information be available. Users should be provided with a clear escape from the conversation. If a user finds that a chatbot is no longer useful, or is incapable of answering their query, there should be a clear means to transfer the conversation to a human agent (if available), or to send email correspondence. Additionally, if a chatbot has answered a query and the user has ended the session or refrained from answering another question, the chatbot should politely end the conversation.

6.1.3. Automated Decision Support

Improving users' experiences when interacting with government services is important, but the benefits of this work are lost if the wait time to receive eligibility decisions on services is too long. Part of service excellence is cutting wait times, and AI can play a role.

To start, AI can be applied to electronic forms – both user-facing and back-end – to help ensure that data entered meets your institution's standard of quality. This modest application can greatly assist your institution's ability to use the data for decision-making later on.

Processing service applications requires that an analyst review application information, verify to see if it is true and believable, and checking if the information that has been submitted meets the program's

⁹ [6] Demeure, Niewiadomski and Pelachaud, "How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment?" *Presence*, October 2011. Link: http://www.mitpressjournals.org/doi/pdf/10.1162/PRES_a_00065

¹⁰ For more on *Eno* and *Pegg* see: <https://www.accountingtoday.com/opinion/the-tech-take-the-genderless-face-of-accounting-bots>

For more on Google's *Assisant*, see: <https://www.engadget.com/2016/10/07/google-assistant-desexualize-ai/>

Responsible Artificial Intelligence in the Government of Canada

eligibility criteria. This process can take time, both due to the amount of information collected as well as the limitations on resources.

By using appropriate program-related input data and a model to test inputs against rules, such as legislative or regulatory requirements, an automated system may be able to process eligibility decisions faster than and as well as a human in many circumstances. This allows eligibility analysis to be processed outside of core work hours, for data analytics to be gleaned and acted upon promptly and organically, and for patterns to be established so that particularly complex or unexpected applications can be investigated more thoroughly. Strictly speaking, this approach can be done without the use of AI, as the rules themselves are strictly defined by the institution.

This level of decision automation has been tested and deployed in private sector settings for over a decade. Insurance and financial sectors have been pioneers in decision automation to improve service response times and to increase fraud detection. These sectors have similar challenges to governments: mission-critical systems with many dependencies, limited budgets and competing priorities for IT development, and a desire to maximize transaction throughput and minimize fraud.¹¹

What if the system was designed in such a way that humans did not choose the eligibility criteria at all, but allowed a machine to determine what applicants should be eligible based on desired outcomes? For example, imagine a hypothetical program that provides small grants to exporters. Rather than have the program experts select the eligibility requirements themselves, an AI system analyzes similar firms in similar industries, and determines the likelihood of success following the grant. Of course, choosing the metrics that define “success” remains the responsibility of the program, but the criteria may vary. Perhaps there are different predictors of success for different sectors, or predictors that human analysts missed.

This approach has the potential to provide services with more effective outcomes, but brings challenges. For example, criteria are often enshrined in legal authorities. If there is a challenge to the decision, the institution would require to show what criteria were used to make the decision, something that might be difficult to show using current technology. This issue is further elaborated [below](#).

Many government services have existed for decades; assuming there is high-quality, machine-readable data available, there is a significant volume of potential training sets to train AI how to process eligibility. By showing AI examples of successful versus unsuccessful applications, it can determine the necessary patterns to extend this reasoning to a new application on its own, effectively mimicking the experience of a human. For this to work, institutions need to have data on the outcomes of services in a format that is readable by machine.

6.1.3.1. Appropriateness of Automation

Should a service be automated completely from end-to-end, or should human intervention and approval always be required? The suitability of an automated system to deliver end-to-end services must be

¹¹ See McKinsey report, “Automating the bank’s back office,” Link: <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/automating-the-banks-back-office>

Responsible Artificial Intelligence in the Government of Canada

analyzed on a case by case basis. Much depends on the type of decision being made and the amount of discretion that any particular decision requires. Departments will have to carefully consider:

- Whether they are acting within lawful boundaries;
- Whether additional authorities are required;
- The procedures and mechanisms to be implemented to ensure transparency and to be able to document how a decision was reached, especially when such a decision affects individual rights and privileges and involves the exercise of discretion.

A “human in the loop” may not be straightforward in overturning machine decisions. Unless they are specifically instructed to, human officers will need to bring themselves to question the authoritativeness of the machine recommendation. Enough information would have to be provided to the human - both from the original input data such as a benefit application - but also around the rationale behind the decision. The human analyst should be required to document why the machine recommendation was not followed. Machine decisions flagged for human approval or overturning would themselves have to be monitored to ensure that there is no internal conspiracy or mismanagement.

The Government of Canada provides a diverse set of programs and services across over 140 federal institutions. Some of these programs and services are critical to the fundamental well-being of people, the economy, and the state; others are less. Should the same rigorous governance and accountability measures be required for non-critical programs as critical ones? Can we classify programs and services into risk categories to better target governance to be proportional to risk?

As TBS prepares guidance on how institutions can responsibly introduce automated decision support to their organization, it will develop a tool by which institutions can assess the degree of automation that is appropriate for their program. Guidance on governance could then be linked to the risk score.

6.1.3.2. Transparency and Recourse

How much information should be provided to users on the decision-making process? The ability and need to explain algorithmic decision making requires a delicate balance. On one hand, transparency builds trust and social acceptance, and provides users with information with which they can challenge decisions and business processes. On the other hand, providing too much information to the public can open a door to malicious manipulation of the algorithm.

Users should be notified in advance of submitting an application that it will be processed by an algorithm, along with a link leading a webpage with accessible, non-technical information on the decision-making process. This information should include a description of the sources of data used to make the decision, and links to recent system performance audits.

Further research is required to determine whether users should be provided an opportunity to opt out of automated decision making in advance of applying for a service. On one hand, this provides users with more control over how their personal information is handled. On the other hand, designing systems for this to occur may be impractical and expensive. Regardless, in the event of a negative decision, users

Responsible Artificial Intelligence in the Government of Canada

should be provided with an opportunity to have their application revisited by an informed human case assessor.

Further research is also required on what information institutions should provide on the design and functionality of AI tools (algorithms, logic, decision making rules), understanding that algorithms may be manipulated with too much of this information.

Regardless of the methodology used, it's important that institutions only automate a process when they have obtained a high level of confidence in the decisions that it is making in a test environment.

6.2. AI to help design policy and respond to risk

What if we were more accurately able to predict migration flows, forest fires, or the impact of an aging population? What if we knew in advance which ports of entry would be more likely to encounter contraband, or which consumer products might be more susceptible to recall? Existing analytical models have already given the GC the ability to better understand certain social or environmental outcomes to policy, but with new methods able to identify patterns in data that perhaps humans were previously incapable of doing, we may be able to make more precise and informed predictions than ever before.

Governments work with big problems. We work in an environment often marked by complex, interdependent systems, where small policy changes can result in massive impacts among a population or the economy. If we can use data to predict the impact of our work with greater precision, or to understand future pressures on social or economic programs, then we can respond more efficiently and ensure that regulatory resources are focused on the highest risk elements of their industries.

Using both structured and unstructured data sources, institutions can enhance their ability to understand what is happening in society and the economy, both in Canada and beyond. This will allow for more effective regulation of industries, as well as more informed policy planning through the use of simulation. The ability to combine even anonymized data sets across institutions in real time may be able to provide policymakers with new insights as to what is causing certain outcomes in society.

There are some limitations to this approach. Predictions are extrapolations of patterns that appeared in the past; while access to vast data sets brings greater opportunity to predict in a complex system, AI can't make truly novel predictions, because the past is not necessarily an indicator of the future. Like all AI systems, the right quantity and quality of data will need to be accessible to make accurate predictions. There is also a risk that predictions are made using data that has been collected in a way that is biased or not fully representative of the world that we live in; this issue is further discussed below.

Already, many federal institutions use a method to describe and compare the degree of risk involved with providing a service to a user. This "risk scoring" technique can be an efficient method to associate an administrative action with risk. To date, this has most often been accomplished using methods that require institutions precisely defining what risk is in their universe. These "closed-rule" algorithms, while not AI, are a form of automation that has shown to be service-enabling by reducing compliance and enforcement burden on lower-risk users.

Responsible Artificial Intelligence in the Government of Canada

6.3. Applying AI to the internal services of government

A professional public service is supported by intuitive and efficient internal services. Some of these services directly service Canada's democratic institutions, such as access to information or responses to the questions of parliamentarians. Others are in place to ensure that the public service itself is functioning smoothly, fostering a positive work environment and securing public assets.

6.3.1. Information Management

From white papers such as this one, to briefing notes, presentations, data sets, and other analysis, the GC is sitting on a vast trove of data, structured and unstructured, tagged and untagged. Traditional means of using this data has been limited to specific, machine-readable formats, but advances in semantic analysis have unlocked the potential for information in text format to be mined for insights as well. Now machine-usable information can be gleaned from text, audio, or video.

This technology can be used for a variety of applications, such as analyzing social media reaction to government policy or events; summarizing past briefings or approaches to maintain institutional memory; or automatically creating documentation trails for internal audit purposes.

The power behind these applications offers the promise of AI eventually providing virtual librarian services. With properly structured and tagged text data, a policy analyst will be able to more easily sort through and summarize past approaches to a problem, or find what is being done in other institutions. Having a smarter content management system understand what an analyst is looking for will help ensure that policy options are driven by data and that corporate memory is retained, leading to greater institutional wisdom.

6.3.2. Automated Content Generation

Over the past several years, products have entered the market allowing for content, be it text, audio, or visual content, to be generated automatically. Systems have been deployed in the private sector to automatically produce newspaper articles, blog content, or marketing copy. One notable example of this technology has been at the Associated Press newswire, which is estimated to be able to generate 2,000 news articles a second. After several months of training, configuration, and maintenance, the system is now able to post stories without any human intervention at all. The "AI journalist" is capable of doing this because a) there was a dataset large enough for the computer to extract best practices, and b) most of these reports contain only factual information, with limited nuance.

There are potential applications for the business of government. This technology can likely be adapted to a number of government documents that are produced on a regular basis in large quantities that are often factual and follow a certain formula or template. While certainly incapable of making normative considerations, this technology can be useful to summarize and compare. For example, it would be able to write Ministerial correspondence, background sections of briefing or meeting scenario notes, background of Question Period notes, etc. This would allow human public servants to focus on analysis, policy lenses, considerations, and strategies for next steps.

Responsible Artificial Intelligence in the Government of Canada

6.3.3. People Management

AI is transforming the discipline of human resources management, whether to gauge and optimize productivity, or to match individuals to suitable jobs. The ability to scan through the information of thousands of candidates using a more precise and insightful method than static keyword searches can potentially lead to more effective hiring decisions. Understanding the skills and credentials of effective and ineffective employees can provide insight as to the attributes of an ideal candidate. This can improve overall organizational effectiveness, but also help an individual find a job they may be ideal for but may lack traditional qualifications.

Another HR application of AI is performance assessment and management. These tools measure an employee's effectiveness against certain criteria, such as delivering on projects or replying to stakeholder inquiries. Using these tools, a manager is able to have a dashboard of the productivity of employees and the current status of their projects.

These tools can bring ethical risks and must be deployed with great care. For many of these systems to work properly, a continuous volume of data must be collected about a person's productivity. This is tantamount to ongoing surveillance of the employee, something that could cause harm to the employee's mental health.¹² Deep and persistent AI supervision of employees may contribute to the very anxiety that reduces their effectiveness at work, which in turn may hinder them from changing jobs. Furthermore, this system would have to reflect the changing context of a job, such as busy or quieter periods of work (i.e. in media relations), or jobs that produce work that is difficult to easily quantify (i.e. policy advice).

Additionally, identifying optimal productivity may fail in certain cultural contexts, as some employees may work differently. A veteran, indigenous person, or someone born abroad may choose to work different hours, or using different techniques, which while effective, may be difficult to measure. An AI trained only on employees of European descent may not effectively evaluate an employee that is not. The systems would be required to consider the diverse accommodations that may be required for employees with certain disabilities.

At the current state of technology, AI systems should be prohibited from making unsupervised decisions about HR. When AI is generating recommendations for management, it is very important that employees be made aware of them in advance if at all possible, and be provided with the opportunity to access the information collected about them.

6.3.4. Security and Access Management

AI can be applied to the way institutions provide, review or revoke IT system and building authorizations by establishing baseline normal behaviour of staff and learning when certain activities seem out of the ordinary. It can provide a better alignment of IT security with operations and reduce the number of ad-hoc requests for access to a system. This can reduce the workload of IT administrators, allowing them to focus on user needs that are exceptional.

¹² <http://onlinelibrary.wiley.com/doi/10.1111/ntwe.12039/abstract>

Responsible Artificial Intelligence in the Government of Canada

AI-powered cybersecurity and access control can further assist by allowing the detection of user needs at a granular level within a very short time, allowing users to have permissions better suited for what their job actually requires. AI can also be used to optimize permissions in business continuity planning.

Finally, there have been advances in machine learning cybersecurity applications that are designed to identify threats earlier, including internal threats where a sudden change of behaviour raises concern. While AI offers great promise in cybersecurity, it should be viewed as a single layer of protection, and not a substitute for existing systems and processes.

6.3.5. Financial Management

Whereas standard data analytics can provide significant value to institutions by helping them understand patterns in their accounting, advances in machine learning and natural language processing have led to a variety of applications for more intelligent financial management.

For example, contract intelligence applications help organizations automate contract review by scanning for mistakes and suggesting corrections. Machine learning systems are also available to help organizations continuously monitor for fraudulent or misappropriated expenditures by learning typical expenditure behaviour in flagging potential anomalies.

7. Policy, Ethical, and Legal Considerations of AI

With all of the potential use cases offering to improve policy and services, enthusiasm for AI in government has been high. Unfortunately, improper application of this technology can lead to negative outcomes for users, from frustrating service experiences to being mistakenly denied eligibility for benefits.

While the use of AI offers a lot of promise in improving the efficiency of government, it is important to approach its use with a strong ethical foundation. Machine ethics have been debated for years, and the Government of Canada should learn from these groundbreaking discussions to ensure that this transformative technology best serves the interest of everyone living in Canada.

As these agents grow to operate in increasingly sophisticated spaces, they act on behalf of the Crown, and should be subject to similar values, ethics, and laws as public servants and adherence to international human rights obligations. Institutions should incorporate these ethical principles in their application of AI:

1. **People should always be governed – and perceive to be governed – by people;**
2. **AI systems deployed on behalf of government should be trained to reflect the *Values and Ethics of the Public Sector* as well as Canadian and international human rights obligations; they should be used to reinforce these values where possible;**

Responsible Artificial Intelligence in the Government of Canada

3. **Organizations are accountable for the actions of AI systems, and should build systems that are auditable;**
4. **Understanding the need to protect privacy and national security, AI systems should be deployed in the most transparent manner possible;**
5. **Organizations should ensure that reliable contingencies are in place for when AI systems fail, or to provide services to those unable to access these systems;**
6. **AI systems should be developed in a diverse team that includes individuals capable of assessing the ethical and socioeconomic implications of the system;**
7. **AI systems should be deployed in a manner that minimizes negative impact to employees where possible, and should, where feasible, be created alongside the employees that will work with them.**

The Government of Canada is committed to incorporating international norms and standards in ethical design when applying AI or any autonomous system. The first step to preventing negative outcomes is to understand what they are and how they occur.

There is no “average” Canadian; this country consists of a population diverse in background and circumstance. There will be users with unique challenges that will test the rigour and limitations of algorithms deployed by government. Institutions need to account for exceptions, minimizing cases that fall through the cracks, and providing recourse for the inevitable failures of the system.

7.1. Data, Bias, and Rights

Every field of data entered is an investment for the future. That data will be examined, validated, and manipulated individually and in aggregate possibly thousands of times in the cycle of their life. Traditionally, data entry was viewed as an input cost to be minimized by many federal institutions, but as the world moves more towards data-driven decisions, organizations are centering data governance in their core operations. This has unfortunately revealed a lack of consistent quality in data holdings.

Many AI applications are only as effective as the quality and quantity of their input data. The first step for an institution wishing to deploy an AI application is to ensure that the necessary training data is available, representative of the problem that needs to be solved, is readable by machine, and that the organization has the legal authority to collect and use this data. It also means adopting a culture of good data practices, and investing in the people and systems necessary to create, store, protect, and use data effectively.

7.1.1. Prevention of Data Bias

AI systems are not neutral; they will learn the biases of its programmers and the datasets used to train it. While unintentional, this bias can have ramifications that could range from embarrassing to serious. Even data that is incorrectly entered or labeled can have knock-on effects that affect real people in real ways.

Responsible Artificial Intelligence in the Government of Canada

This can particularly affect vulnerable populations, of whom data has been collected historically with varying quantity and quality.

The ability to distinguish, predict, and learn means that AI is able to operate in a more abstract and probabilistic fashion than earlier forms of computing. To do this, AI needs to be trained with datasets and oriented towards preferable outcomes. Both the training process and the selection of preferable outcomes carry with it the bias of the humans that collected and tagged the data, as well as the programmers that designed the algorithm. The collection of some data can be imperfect due to social or cultural stigma; for example, suicides and sexual assaults in Canada are both underreported.^{13,14} Even the choice of which datasets to use and which to reject may entrench bias into the decision, and can lead to different outcomes.

Without enough training, an AI will have difficulty achieving its task, or will do so in a way that could lead to misinterpretations of data. Data collected in a certain socioeconomic context will echo in the decision-making of algorithms. The responsible policy manager needs to ensure that this important context is added to the analysis, and that they understand potential ways that AI can interpret input data incorrectly. Even controlling for certain variables won't necessarily protect from bias, as it can be derived from other, correlated variables; for example, excluding ethnicity from analysis won't necessarily protect from bias if the system can infer ethnicity from another variable such as a name.

The results of data bias can be highly problematic. As AI applications are more widely dispersed throughout society, a number of these unintentional but notable biases have been uncovered. For example, an algorithm used to predict crime in the United States has been shown to reinforce discriminatory policing because the crime data upon which it was trained was collected disproportionately in African-American neighbourhoods.¹⁵ According to a study by Carnegie Mellon University,¹⁶ women tend to be shown job ads for high-paying jobs less often than men as a result of search algorithms, likely due to the fact that women have been disproportionately missing from these positions in the past.

Machines can't learn contextual policy objectives such as social equity or environmental stewardship without being taught that these goals – while maybe not the explicit goal of the system – are necessary trajectories to be taken into consideration.

Algorithms themselves can affect the systems that they are trying to assess through a feedback loop. For example, a recidivism model that determines early release from prison, but being in prison longer increases the probability of recidivism, creating a feedback loop that increases incarceration time.

In applications where machine vision or audition may be applied to individuals, it's important that people are not excluded by virtue of ethnicity, accent, or disability. Some rare disabilities may not appear in

¹³ <https://www.statcan.gc.ca/pub/82-624-x/2012001/article/11696-eng.htm>

¹⁴ <https://www.statcan.gc.ca/pub/85-002-x/2017001/article/14842-eng.htm>

¹⁵ <https://mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows#.sGlb3QeCM>

¹⁶ 2015 study using 1,000 simulated persons, Link: <http://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html>

Responsible Artificial Intelligence in the Government of Canada

training data at all, which could lead to negative outcomes for individual. For example, a border camera scanning for predictors of risk may misinterpret a “tic” of an individual with Tourette syndrome as suspicious. These can manifest in a diverse fashion, and should not cause this person to undergo secondary inspection every time they pass through the border.

Social media presents an unprecedented opportunity to understand what some Canadians happen to be talking about on some subjects, but this approach brings significant risks. First, many Canadians do not use social media, so views cannot necessarily be taken as representative of the population. For example, in 2016, only 22% of Canadians used Twitter.¹⁷ Second, national government or private social media firms are capable of marshalling thousands of social media accounts that are artificial, expressing whatever views that they are paid to express. These botnet campaigns can inflate the prevalence of certain perspectives. Without strong countermeasures to detect deliberate attempts to distort public discourse with botnets, social media data should never be assumed to have been produced exclusively by humans.

7.1.2. Data for Insights and Privacy Rights

While many of the privacy risks brought by AI are not fundamentally new, the magnitude of data collected and the ability to manipulate this data beyond what a human is capable of brings a new dimension to these risks.

Algorithms capable of gathering insights from unstructured data mean that the nature of the information that we collect changes. For example, a name is no longer just a name, but a data point in a wider pattern that can reveal ethnic background. The government did not explicitly collect ethnicity, but an algorithm could extract that information from the person’s name nonetheless. One’s address can reveal correlations with income, health outcomes, or likelihood to encounter crime. With only a few variables known about an individual, it is possible to extrapolate an entire portrait about a person that is potentially very personal, and surely more than the person intended to disclose. This extrapolated information is not verified to be true; it is simply resulting from a series of statistical correlations that imply it may be true.

This can have unintended consequences. As an illustrative example, suppose that an institution wanted to predict the most successful outcomes possible from a grant program. It trains an algorithm based on a variety of historic information on the companies that typically apply, their officers, and the outcomes of those grants several years, themselves determined using public sources. Surprisingly, the findings are that women who have undergone a change of name are at a high risk of their business is failing.

Collecting – or rather, deriving – this new personal information from the individual is not necessarily unethical; the test lies in how this insight is used. In this case, the institution should compensate for the bias to ensure that these women have an equal opportunity to receive the grant.

To the extent that AI uses personal information it must comply with the *Privacy Act* or other departmental privacy codes. For the purposes of the *Privacy Act* and Privacy Impact Assessments (PIA), AI does not

¹⁷ <http://www.digitalnewsreport.org/survey/2017/canada-2017/>

Responsible Artificial Intelligence in the Government of Canada

represent a new program, but a new suite of tools. Existing program PIAs should be updated to reflect these new tools, as well as the new data that may be collected and used for the program.

For the purposes of Privacy Impact Assessments (PIA), AI does not represent a new program, but a new suite of tools. Existing program PIAs should be updated to reflect these new tools, as well as the new data that may be collected and used for the program.

7.2. Transparency and Accountability

7.2.1. Accounting for the Actions of AI: The “Black box” Problem

A cornerstone of responsible government is that Ministers are accountable for the affairs of their portfolio institutions, enshrined in legislation and custom. This accountability flows down from them to Deputy Heads and others within institutions through a variety of authorities from law to regulation to Treasury Board policy.

If a human makes a decision that is challenged, we rely on his or her explanation, the data that they were exposed to, and the outcome to figure out what happened. The limitations to this approach reflect the limitations of humans; memory may be unreliable, notes can be incomplete, and human biases can affect recall. Society relies on human judgement partially because it is the only option available to us.

Recall that advanced machine learning methods used today, such as neural networks, involve breaking down a problem into thousands – if not millions – of small decisions in order to reach an outcome. Some of these decisions are explicitly coded into the algorithm, but others are learned by the algorithm on its own based on input data. Much like the human brain, it is difficult to understand the entire decision-making process in detail made by the machine’s artificial neural network. This is known as the “black box” problem.

If an algorithm needs to be examined for whatever reason to determine the decision making process, hundreds of thousands of lines of code may need to be reviewed. Even then, it might be very difficult to reproduce exact results, or determine exactly why an outcome occurred. In the context of a neural network, examining each artificial neuron may not provide a sufficient understanding of the decision-making process.¹⁸

Invalid or biased decisions by algorithms tend to exist due to incomplete or biased datasets; therefore institutions should focus on the quality and completeness of training data, testing and audit findings of the applications, and any operating parameters. Systems should be continually tested and audited to ensure that the outputs still meet the original intention.

¹⁸ This is a rapidly-evolving area of research. Statistical and cryptographic techniques (e.g. Merkle trees) have been suggested to resolve this problem by creating an audit trail of decisions

Responsible Artificial Intelligence in the Government of Canada

Technology may be able to solve some of these problems; for example, there are tools in development¹⁹ that can trace and describe how a decision was made using a neural network.²⁰ TBS will have to continually monitor the development of this technology to ensure that transparency-enhancing techniques are adopted. Ultimately, how much explainability is required will largely be determined by jurisprudence. It is expected that this will be a higher test than what is expected from a human case manager. That said, institutions must ensure that decision-making algorithms provide enough detail around an explanation to understand why it was made for reasons of administrative and legal oversight like that carried out by the Information Commissioner, the Privacy Commissioner, the Auditor General, and the Courts.

If the government has to make decisions based on models that they don't understand or have access to, it hands some decision-making power to a private company with a black box. It is important that institutions have full understanding of the tools that they are using for decision support. To manage this risk, institutions may have to develop algorithms using internal or contracted resources, and maintain ownership of the input data and the algorithms used to make decisions. As well, institutions will have to retain all data used to train an AI for the duration that the AI is in use.

7.2.2. Social Acceptability

The enduring nature of Canadian democracy rests in the provision of good government to its citizens and residents. The Government of Canada is exploring the use of powerful tools at a time when trust in public institutions is low,²¹ and when a minority of Canadians feel that new technologies will do more good than harm.²² Not only do government AI applications need to be effective, but the population needs to perceive them as effective as well for them to be legitimate additions to, or substitutes for, human officials. If AI is going to make decisions, recommendations, or help design policy, there needs to be a sufficient level of social trust that these systems work, and work to the population's benefit. If the trust and support does not exist, then these tools will fail.

Trust will be built over time, assuming that the rules surrounding the use of these tools are transparent, and that appropriate information is available to users about how they work.

7.3. AI and the Law: An Emerging Landscape

AI will likely challenge current legal paradigms, although its actual impact on the law is still uncertain. In Canada, the technology is not comprehensively regulated, and there are few cases involving AI that have gone to court. If a government institution implements an AI solution that collects, uses, discloses or

¹⁹ <https://qz.com/1022156/mit-researchers-can-now-track-artificial-intelligences-decisions-back-to-single-neurons/>

²⁰ See for example: Pang Wei Koh, Percy Liang; Proceedings of the 34th International Conference on Machine Learning, PMLR 70:1885-1894, 2017 - Link: <http://proceedings.mlr.press/v70/koh17a.html>

²¹ <https://www.edelman.com/trust2017/trust-in-canada/>

²² Ipsos Canada Next, Public Perspectives. October 2017. <https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-10/public-perspectives-canadanext-2017-10-v1.pdf>

Responsible Artificial Intelligence in the Government of Canada

retains personal information various requirements of the federal *Privacy Act* may come into play as well as additional requirements found in applicable program or departmental statutes.

The use of AI in the government will undoubtedly have legal implications that range across many diverse areas of law, including Administrative Law, Privacy Law, Cyber-security Law, Intellectual Property Law, Crown Liability, Charter and Human Rights Law, Procurement Law, Employment Law, and the Law of Evidence.

Legal issues will be raised at each stage of use of AI, from its development to its deployment. For this reason, it will be very important to ensure that its use protects people's fundamental rights and that ethical and legal standards are considered at each stage of the use of AI, from the earliest stages of development onwards. To this effect, institutions should engage their institution's legal services unit as early as possible in the design and development of their project.

7.4. Technical Considerations

With the rapid expansion of AI-driven applications in the last five years, there have been moves to ensure that technical standards are established for interoperability, best practices, and increasingly, safety as well. International standard-setting organizations such as the IEEE Standards Association, OpenAI, and ISO have established technical standard working groups to bring to the tools, methods, and practices associated with algorithm development and implementation. With an aim to adopting international best practices, TBS will be closely following the developments of these standards with an aim to provide guidance to institutions soon as appropriate.

Analytics and AI projects are inherently multidisciplinary and therefore will cross many teams and branches in an organization. Federal institutions exploring this technology should be sure to include their IT and data management teams from the design stage onward. These applications will need to fit consistently in the enterprise architecture to allow for secure connectivity with client relationship management software and other data repositories both within your organization and beyond. Chatbots, for example, may need to interact with your web presence if they include navigational capability. While the communications or policy function within an institution may be the business owners of this technology, it's important to remember that there may need to be connections with an institutions client relationship management software, or need to draw data from other organizations.

Institutions will need access to the right tools in a timely fashion. Data science teams should have access to the required software and servers without undue delay. Access to secure cloud computing services rated to protected B will also be required.

7.4.1. Cybersecurity considerations

Many of the cyber security implications to AI are similar to those of other critical systems and government. However, there are some new threats to take into consideration.

Responsible Artificial Intelligence in the Government of Canada

Databases containing training data must be properly secured from intrusion. Even if the intruder cannot extract meaning from the data, changing figures in training data can lead to changing the outputs of the algorithms. The effects of these changes could be wide-ranging, unanticipated, and difficult to detect.

As mentioned above, deep learning chatbots can be taught to provide incorrect or inappropriate information. Keyword blacklists only go so far to prevent inappropriate behavior, because a chatbot doesn't necessarily need to use explicitly bad language to provide bad service. The training data provided to chat box need to be properly secured, so that an intruder cannot, for example, reroute a link from a government service to of malicious clone used to steal personal information, or be tricked into behaving in inappropriate ways.

An intruder may want to shut down an automated decision support system that an institution relies upon. Like many critical systems, redundancies will need to be put in place to prevent lengthy outages to critical systems. This should include that at least some human staff are retained and properly trained for manual backups.

8. Rethinking a Post-AI Enterprise

AI tools will be transformative for government, but only if institutions are ready to deploy them. The first step is investing in data science and business intelligence capacity within your Institution. This includes skilled personnel, tools, storage, and data governance mechanisms. These investments should be led by a skilled Chief Data Officer, who has control and access to data sets throughout the organization, an understanding of related data holdings of others, and a direct reporting relationship to the Chief Information Officer of the organization or a suitable Assistant Deputy Minister.

The second step is to ensure that the deployment of AI applications come with a multidisciplinary and diverse project team. Having a mix of social scientists, ethicists, data scientists, change management professionals and user experience designers from a variety of backgrounds is a potent defense against data biases and other risks preventing your organization from reaching success. Easy and accessible user experience is vital to the uptake of these tools, regardless if they are not directly used by the public.

8.1. New Approaches to the Workforce

Advances in artificial intelligence have brought with them a very public discussion about the future of work and the role that knowledge workers will play in the economy. To date, there are examples of task automation in the private sector that lead to significant staff reductions; conversely, there are examples where no workforce reductions were required. AI has shown itself to be a valuable suite of tools that can either exponentially increase productivity, or eliminate routine tasks, allowing for humans to perform more valuable work for the organization. That said, this transition does not happen on its own, and anxiety that can result from perceived imminent automation can have a real impact on employee

Responsible Artificial Intelligence in the Government of Canada

wellbeing and productivity.²³ As institutions look to automate work, it is important that they choose approaches that maximize utility to the organization while at the same time minimizing the potential for staff reductions.

At the same time, post-automation and AI government will require new skills from existing staff and new types of staff from the labour market. Federal institutions will need to attract data scientists, invest in up-to-date tools that they can use, and access to relevant data sources.

Institutions will need to ensure that their policy analysis and development teams understand how to access, interpret, and manipulate data relevant to their work, and to have access to the skills development resources that they need to grow within their field. New employees to a team should be provided with context on how their data has been collected and used in the past, and how stakeholders typically view this collection and use.

Federal institutions should also be reminded that some collective bargaining agreements contain specific sections on workforce adjustment due to technological change. To ensure that these requirements are honored, TBS recommends that unions and non-represented staff alike are engaged early in the planning phases. Staff and unions will be useful partners to help automate processes in a way that is both most useful to the user as well as least affecting of positions.

8.2. Evolving How Government Works

Departmental internal audit, central agencies, and agents of Parliament together play a role to ensure that programs are designed in a manner that is compliant with policy and best practices. This robust system of oversight ensures that institutions are accountable to ministers, Parliament, and the public. In an era of increasingly data-driven policy recommendations and autonomous systems, does the government have the right tools to oversee its business?

There is no organization that currently exists with the clear mandate and capability to respond to complaints of data biases or algorithmic design. Canadians and parliamentarians alike will need an obvious contact point to manage these issues, and oversight organizations provided the tools necessary to do this job. While algorithms cannot be completely transparent to avoid fraud, an oversight body staffed with the required expertise could be provided access when required.

TBS will require to do further research on models of governance that could provide the necessary oversight and guidance to Federal institutions. This can range from an *ad hoc* federal “Automation Advisory Board” comprising of internal and external experts to a more formal and permanent body with staff. Regardless of the model chosen, the body would have the ability to review automated decision-making by any methodology, and provide advice to ministers during the design - but especially prior to Cabinet approval of projects - on ethical design of AI-driven programs and services.

²³ The Economist, “Automation and Anxiety,” June 2016. Link: <http://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety>

Responsible Artificial Intelligence in the Government of Canada

9. Conclusion

AI applications have emerged as useful tools for institutions to include in their policy, program, and service development process. They can bring exponential power to government, but must be applied where it makes sense. While AI applications can be applied to many projects, but should only be considered if there is a reasonable value proposition from its use. Simply adding a machine learning component to a project will neither guarantee its success nor be the sole guarantor of smarter policymaking. Introducing AI into a program brings risks that need managing, and requires staff capable of managing it.

AI's complex and multidisciplinary nature demands that federal institutions work together, sharing talent and best practices, to leverage knowledge and avoid duplication. It means working with other orders of government, and with research institutions and non-governmental organizations within Canada and beyond.

Government is also fundamentally about people and relationships. Machines cannot substitute for empathy, and even the best analytics will find outliers that cannot be forgotten. TBS actively encourages institutions to explore this technology for the benefit of the populations that we serve. Ethical and responsible design of these systems will drive a virtuous cycle of acceptance, which in turn will drive further development.

As a next step, TBS will begin to examine its policy suite to ensure that existing guidance is useful to institutions that will implement AI applications in their organizations. Where appropriate, standalone guidance will be considered as well. As a rapidly evolving technology, TBS will require ongoing engagement to ensure that policy is reflective of technological capabilities so that institutions can continually make best use of what AI has to offer.

Responsible Artificial Intelligence in the Government of Canada

10. Acknowledgements

This paper was drafted using an “open” approach, with contributors welcome to engage in its development from conception to completion. Treasury Board of Canada Secretariat would like to thank the numerous and ongoing contributions from participants in all sectors.

Responsible Artificial Intelligence in the Government of Canada

[8] <http://onlinelibrary.wiley.com/doi/10.1111/ntwe.12039/abstract>

[16] <http://www.ekospolitics.com/index.php/2014/01/looking-back-and-looking-forward-part-2/>

Will contributors be thanked in this version?

ALGORITHMIC IMPACT ASSESSMENT (AIA)

INTRODUCTION

Governments are increasingly looking to utilize Automated Decision Systems to make, or assist in making administrative decisions. However, questions are being raised as to how to adequately assess the impacts that these systems may have on Canadians and federal institutions. The aim of the proposed Algorithmic Impact Assessment is to develop a framework to help institutions better understand and mitigate the risks associated with Automated Decision Systems and to provide the appropriate governance, oversight and reporting/audit requirements that best match the type of application being designed. The AIA supports the [TBS Standard on Automated Decision-Making](#).

In developing this proposed AIA Framework, the following three key goals have been identified:

- Increase capacity to evaluate the impact of Automated Decision Systems including legal and ethical issues, such as disparate impacts and due process 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.

APPROACH

This document intends to set the broad categories and key questions that will need to be considered in an AIA. The objective will be to use these questions to then build an electronic application that uses decision trees (or other appropriate techniques) to provide institutions with a “Risk Score” based on the type of Automated Decision System they are considering. Each category of risk will provide different recommendations and considerations.

LIST OF QUESTIONS

A) ESTABLISHING SCOPE AND PROJECT DESIGN

Business Requirements

- Will the decision made or recommended by the system involve eligibility determination on the following: (Check all that apply)
 - Restricting movement into, out of, or within Canada
 - Granting access to a private premises or network

- Funding to an individual, business, or other organization
 - Issuing a fine, administrative monetary penalty, or collection notice
 - Issuing licenses or permits
 - Restructuring a market, sector, or segment thereof
 - Granting status designation
- What are the key business drivers for automating the administrative decision-making process? (Check all that apply)
 - Existing backlog of work or cases
 - Improve overall quality of decisions
 - Lower transaction costs of an existing program
 - The system is performing tasks that humans could not accomplish in a reasonable period of time
 - Improve client service
 - (Other)

Development Considerations

- Is the scope of the system clearly reflected in project documentation?
- Has the design team consulted the following experts within their organization : (Check all that apply)
 - Communications
 - Data and Information Management
 - Enterprise Architecture
 - Human Resources
 - IT/Cybersecurity
 - Legal
 - Service Delivery
 - Strategic Policy
- Have you developed Key Performance Indicators and benchmarks to measure and improve the system's performance?
- Is there a process in place to document how data quality issues were resolved during the design process?
- Have appropriate strategies been developed to manage the risk that outdated or unreliable data is used to make an automated decision?

GBA+ Analysis

- Have you undertaken [Gender Based Analysis Plus](#) of the training data?
- If so, did the analysis highlight any risk to a particular gender, gender expression or identity, or visible minority?

- If so, have you documented mitigation strategies?

Decisions about Indigenous Peoples

- Will your system make decisions specifically about Indigenous peoples or territory?
- If so, did you collect your data in a manner adherent to OCAP™ principles
<http://fnigc.ca/ocapr.html>?
- Have you engaged Indigenous organizations in meaningful consultation in the design of the system?
- If so, which ones? (Check all that apply)
 - Assembly of First Nations
 - Congress of Aboriginal Peoples
 - First Nations Information Governance Centre
 - Inuit Tapiriit Kanatami
 - Métis National Council
 - Native Women's Association of Canada
 - First Nations Technology Council
 - Band councils or Indigenous governments in relevant or affected regions
 - Other: (_____)

Machine Vision

- Does your system use contactless machine vision for biometric recognition (e.g. facial, palm print, full-body person, gait)?
- If so, does your system perform within acceptable parameters for all skin colours as defined by the Fitzpatrick Skin Type scale or similar measurement?
- Where applicable, have you tested to ensure that the system performs within acceptable parameters with persons who have a disability that may cause them to be unduly disadvantaged by the machine vision component? (For example, ensuring that a gait analysis tool does not unduly disadvantage an individual that uses a wheelchair)

Governance

- Does the system have a governance structure?
- Does the governance structure assign accountability for the : (check all that apply)
 - design;
 - development;
 - implementation; and
 - Improvement and maintenance of the system?

Integration with other Systems

- Does the system interface with other IT systems?
- Where it interfaces with other systems, have you clearly identified the business processes that occur between systems?

B) PUBLIC NOTICES AND TRANSPARENCYPublic Notice

- Did you develop a notice indicating that the decision rendered will be undertaken in whole or in part by an Automated Decision System?
- Is that Notice publicly available?
- Was that Notice provided to users at the earliest possible stage of the administrative process?

System Functionality

- Are you planning to make information about system functionality publicly available?
(Check that all applies)
 - Requirement specifications;
 - Training data sources;
 - Decision trees;
 - Confidence thresholds;
 - Pre-defined models
 - Criteria
 - Classification structures.

Explainability

- Does the system provide timely and comprehensive explanations to users as to how the decision was made in their specific case? (Check all that apply)
 - Contributing variables;
 - Decision rules;
 - Reference groups; and
 - Profile groups.

C) ASSESSMENTS: FAIRNESS, JUSTICE, DUE PROCESS AND DISPARATE IMPACTSAdministrative Law Considerations

- Have the relevant areas of legislation been identified during the scoping phase?
- Does the system require the exercise of discretion or judgment by an officer? (i.e. human in the loop)
- Have all the decision points that involve the exercise of discretion or judgement been clearly identified as requiring human input?

- Are the business rules relating to discretion or judgment contained in the automated system open to internal and external review?
- Have the system design and processes been reviewed by the appropriate Legal Services Unit and other legal experts, as required?
 - Has a formal legal opinion been issued?

Privacy and Security Considerations

- Does the Automated Decision System require the use of personal information?
 - If yes, have you conducted a Privacy Impact Assessment?
- Have you conducted a Security Assessment and Authorization? (Yes/no/not applicable)
- Is there a process in place to monitor and grant access permission?

Interventions and Overrides

- Does the system enable human override of system decisions?
- Does the system enable override of human decisions?
- Does the system enable the override of other automated decision systems?
- Are officers able to make manual decisions if necessary?
- Do you have processes in place to report and describe the instances when overrides were performed?

Testing

- Do you have a process in place to test and verify business rules?
- Did you develop verification processes, for example testing for “unintentional outcomes”?
- Did you develop a verification strategy to assess whether the system is being used as designed and intended?

Training

- Is the system used by a different part of the organization than the ones who developed it?
- Did you develop training programs to ensure that the system is used effectively and properly?
- Do you have mechanisms in place to ensure that training requirements are being met?

D) MEANINGFUL ACCESS: RESEARCHERS AND AUDITORS TO REVIEW SYSTEMS

Audit

- Does the audit trail clearly set out all decision points made by the system?

- Can the decisions made by the system be incorporated into a statement, reasons or other written notification, where required?
- Can the system provide an audit trail that can monitor recommendations, decisions and processes?
- Have you designed the audit trail to include clearly identifiable links to: authorized delegations, including relevant legislation, policy or procedures?
- Have you included change control processes in the audit trail to:
 - record modifications to the system's operation or performance; and
 - reflect changes to the legislation that underpins the operation of the system.
- Does the audit trail show who an authorized decision-maker is?

Peer review

- Have you sought the review of the system by a party external to your institution, including relevant algorithms and data sets?
- If so, who? (Check all that apply)
 - Researcher(s) from National Research Council of Canada
 - Researcher(s) from Statistics Canada
 - Qualified individual(s) in another federal institution
 - Qualified individual(s) in another order of government
 - Member(s) of faculty of a post-secondary institution
 - A relevant non-governmental organization
 - Private firm or consulting group
- Do you plan on publishing the system's design documentation and algorithms in a peer reviewed journal?

E) CONSULTATIONS AND CHANGE MANAGEMENT

Communications and Change Management Strategies

- If the system is intended to undertake tasks currently undertaken by human staff, have you engaged your departmental human resources?
- Have you developed a communications strategy to address any potential changes to work practices for officers?
- Have you developed an external communications plan?

User feedback

- Have you established appropriate user/client feedback mechanisms?