Institute on Governance
LEADING EXPERTISE
Institut sur la gouvernance
EXPERTISE DE POINTE

Artificial Intelligence and Canadian Government Science: A Discussion Paper

Marija Cvetkova and Rhonda Moore, September 2023

TABLE OF CONTENTS

ACKOWLEDEMENTS

The IOG acknowledges the contributions of the following individuals who reviewed this paper.

- Benoit Deshaies, Director, Responsible Data and Artificial Intelligence, Office of the Chief Information Officer, Treasury Board of Canada Secretariat
- Stephane Gagnon, Vice-President, Communications Research Centre Canada, Innovation, Science and Economic Development
- Cherif Ould Hamou, Manager, Science and Innovation, Public Services and Procurement Canada
- Justin Schonfeld, Research Scientist, National Microbiology Laboratory, Public Health Agency of Canada, Government of Canada
- Luisa Wang, Director, Strategic Policy Division, Consumer and Hazardous Products Safety Directorate, Healthy Environments and Consumer Safety Branch, Health Canada
- Christina Young, Senior Engineer; Risk Assessment Division, Consumer and Hazardous Products Safety Directorate, Health Canada

INTRODUCTION

In recent years, artificial intelligence (AI) has progressed rapidly in both the sophistication and integration into our everyday lives (ISED, 2022). From tools that verify spelling and grammar to ones that protect our homes, applications of AI seem to be everywhere and all at once. According to KPMG (2023), in 2023, 37% of Canadian businesses were using AI, while most citizens interact with some form of AI daily. Likewise, governmental departments and agencies use the technology to ease administrative processes and improve efficiency in service delivery (Zuiderwijk, 2021; Treasury Board, 2023a). AI systems are also increasingly deployed in science and engineering fields (CCA, 2022). Around the world, states and industries are investing heavily in AI R&D to establish AI strategies to secure their leadership in this fast-developing space (Donahoe, 2019; Smuha, 2021). As Chow and Perrigo (2023) remark, the "AI race [is] changing everything." Yet, AI is not without some genuine risks. Issues of bias, transparency, privacy, trust, and accountability raise concerns for this technology's responsible and ethical deployment.

This paper begins with a primer on AI, including an introduction to key definitions and concepts, the state of AI in Canada, and an examination of key opportunities and risks of using AI technologies. The second part of the paper presents a short summary and discussion of legal and policy frameworks that guide the use of AI for Science Based Departments and Agencies (SBDAs). To complement the information derived from the literature, the IOG prepared and disseminated a short survey to members of the federal scientific and research community to collect information about current uses and applications of AI in science-based departments and agencies. The survey results are presented in part III. Part IV contains a discussion of Canada's AI policy landscape as it specifically relates to the conduct of government science. The paper concludes with some questions for further discussion and research. An account of all technical components referred to throughout the paper can be found in the glossary of terms.

SCOPE

This paper examines the current federal framework guiding the use of AI in Canada and the challenges specific to the federal scientific community. This paper was developed for the ninth cohort of the IOG's Leadership Development Program in Science and Innovation (LDPSI). LDPSI targets public servants in policy or science who aspire to join the Executive ranks of the federal public service. Considering the intended audience, this paper focuses on those legal and policy frameworks which have relevance to federal public servants whose work or mandate is based in science. As such, this paper does not examine provincial or international AI policies, nor does it examine the state of AI in other jurisdictions except for where select examples serve to chronicle an important concept or idea.

METHODOLOGY

The literature review for this paper was conducted in May and June 2023 and focuses on scholarly publications, grey literature, and media sources within the last 5 years which consider and examine AI as it relates to government and government science. The review focuses on publications which explain key definitions of technologies that fall under the AI umbrella, those which explain the opportunities and risks of AI use, and those which consider and analyze regulatory, policy, or legal frameworks relating to the use of AI in Canada. Because much of the literature on AI is of a technical nature, with less robust scholarship published on topics related to governance, regulation, or AI policy, grey literature in the form of government publications, conference papers, lectures, and reports by non-profit organizations supports much of this review.

To further inform this paper of the level of AI activity inside federal science-based departments and agencies, the IOG disseminated a survey to members of the federal research and science community via email. The survey collected responses from 49 individuals between 11 and 25 July 2023. Email replies to the IOG confirm the survey was distributed widely across at least seven federal departments and agencies but no hard data on the home departments of respondents was collected. The survey was designed to protect the confidentiality of respondents. Additional research is required to understand the degree to which AI tools and applications have been integrated into the daily work of federal science-based departments and agencies.

PART I: PRIMER ON AI

DEFINING AI

There is no universally agreed-upon definition of AI. AI is often used in reference to both the software capabilities of machines, systems, and/or the field of research itself (Brookfield, 2018). In fact, Smuha (2021) notes that “there are as many definitions of AI as people talking about it” (p. 62).

In Canada, the Treasury Board Secretariat defines AI as “Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours or solving problems” (Treasury Board, 2023a). The use of the term ‘ordinarily’ in Treasury Board’s definition of AI is worth noting; it may imply that as technology becomes more sophisticated, our concept of what constitutes AI may also evolve. This is known as the “AI effect”, in which every new milestone of AI essentially redefines AI (Haenlein & Kaplan, 2019; Smuha, 2021).

The Government of Canada further defines an automated decision system as “any technology that either assists or replaces the judgment of human decision-makers through rules-based system, regression analysis, predictive analytics, machine learning, deep learning, a neural network or other technique” (Treasury Board, 2023a). An automated decision system could be one that uses AI, amongst other tools or techniques, but not all automated decision systems necessarily employ AI.

KEY CONCEPTS: FORMS OF AI, AI TECHNIQUES AND APPROACES

According to Brookfield (2018), AI is oftentimes conceptualized into 3 distinct capability-based classifications: Narrow AI, Artificial General Intelligence (AGI), and Artificial Super-Intelligence (ASI).

Narrow AI are those systems which apply AI only to specific tasks (Kaplan and Haenlein, 2019). For example, facial recognition software has the capability to perform image or object recognition, but it cannot interpret context or ascribe meaning to the images. Siri has the capability for voice recognition or to set your morning alarm, but cannot autonomously solve problems in other areas, such as driving a car (Kaplan and Haenlein, 2019). Artificial general intelligence (AGI) is generally understood as a system which can perform a wider variety of intellectual tasks and which can simultaneously switch from one task to another, adapt to environmental challenges, and solve new problems at an intelligence level comparable to humans (Allen and West, 2020). Artificial Super-Intelligence, though harder to define, can be understood as a system which can perform all tasks in all areas, outperforming even the smartest humans (Kaplan and Haenlein, 2019).

As of 2023, we have only achieved Narrow AI (Manheim and Kaplan, 2019; Brookfield Institute, 2018) and it is still unclear as to how long it will take to achieve general AI (Francois, 2023), nor can AI experts agree on a timeline for the realization of artificial super-intelligence. Experts in the field have been predicting a human-level AI since the 1960s (Munk Debates, 2023). However, the growing functions of generative AI are promising developments that indicate AI systems could indeed achieve a level artificial general intelligence (AGI) in the very near future. Yoshua Bengio and Geoffrey Hinton—respected researchers in the field of AI—have forecasted that super intelligence is 5 years to decades away (Munk Debates, 2023).

As a research discipline, AI encompasses several sub-fields, or what some have referred to as AI applications, approaches, and/or techniques. The core fields loosely include Machine Learning (ML) and its own subset Deep Learning (DL), Natural Language Processing (NLP), Robotics, and Computer Vision (Kuziemski and Misuraca, 2020; Faculty of Computer Science, 2023). Other streams of AI research exist and often overlap, such as 'Predictive Analytics' (Brookfield, 2018), 'Learning Systems,' and 'Knowledge Representation' (Manheim & Kaplan, 2019; Faculty of Computer Science, 2023). Algorithmic core fields, such as ML and DL, are often employed in application-oriented core fields, such as NLP or Predictive Analytics. Machine Learning and Deep Learning have become the most popular and promising fields because of their far-reaching applications, ranging from fraud detection to self-driving cars.

Machine Learning (ML) is a technique "that enables computer systems to learn and make predictions based on historical data…powered by a machine learning algorithm…that is able to improve its performance over time by training itself using methods of data analysis and analytical modelling" (Brookfield, 2018, p. 4).

There are two primary forms of ML: Statistical ML and Model-driven ML (King & Zenil, 2023).

Statistical ML is the most dominant and successful form of ML and is based on complex pattern learning and powerful statistical computation. Statistical ML works by finding regularities in datasets, which "can then be interpreted or studied further" (King & Zenil, 2023, p. 181). However, large amounts of data are needed for statistical ML approaches, which may not always be available in science. In addition, statistical ML performs poorly in the areas of abstract modelling or logical inferencing.

Model-driven ML: "generated mechanistic models from the data consistent with the data themselves that can be tested against newly generated data" (King & Zenil, 2023, p. 181). The models are mechanistic in the sense that "they can be followed state by state, as in a dynamic system, through a chain of cause and effect" (p. 181). Contrary to statistical ML, model-driven ML does not require robust training data to explain observations, making it a more powerful tool for generating abstract models or generalizations (transfer learning).

Deep Learning (DL) is a stronger form of statistical machine learning (King & Zenil, 2023) that "uses learning algorithms called artificial networks that are loosely inspired by the structure of the human brain. Artificial neurons are connected to one another in layers that rewire and edit themselves on the fly…emulating neural pathways in the brain which strengthen themselves each time they are used" (Manheim & Kaplan, 2019, p. 115). This approach "allows DL to find patterns in unstructured data, from which it models knowledge representation in manner that resembles reasoning" (Manheim & Kaplan, 2019, p. 115).

The modern approach to AI primarily uses Machine Learning and Deep learning "to identify patterns, produce insights, enhance knowledge-based work, and automate routine tasks" (Brookfield Institute, 2018, p. 4). DL constitutes the fastest-growing approach to AI and is what many scholars and experts refer to when they discuss AI. As such, this review's discussion of AI emphasizes DL approaches when discussing AI applications, risks, and opportunities.

ML and DL models are made possible through the use of training data. There are a handful of learning techniques that can be employed to train the models: supervised learning, semi-supervised learning, reinforcement learning, and unsupervised learning. It is important to note that neural networks and neural-net based ML algorithms have pre-dated DL by decades, although the key distinction is a DL model's ability to deal with largely unstructured data with little to no human supervision.

OPPORTUNITIES AND APPLICATIONS IN SCIENCE, GOVERNMENT, AND GOVERNMENT SCIENCE

AI presents tremendous opportunities to solve complex problems across all sectors of the economy. AI systems are already being deployed in the health sector to detect diseases with greater precision and speed, improve treatment plans, and monitor patients (CSPS, 2021a). AI has also contributed to novel scientific discoveries. In 2020, Google's sister company DeepMind solved a 50-year-old biological puzzle: predicting the 3D structure of proteins (Callaway, 2020). AI technology has been used to improve precision harvesting in agriculture and streamline energy supply chains (ISED, 2023). City-planners in metropolises such as Toronto and Montreal are exploring the use of AI in transit systems and building developments, while schools and universities are exploring the use of AI to evaluate applications and improve curricula delivery through 'smart tutoring' (CSPS, 2021a).

In science, AI has been deployed in "basic, applied, and experimental development research" for several years (CCA, 2022, p. 10). AI can uncover pathways or relationships not considered by humans and is able to sift through large unstructured datasets while considering multiple variables. The OECD notes that AI can also assist in uncovering new scientific insights from old literature, optimize scientific workflows, and compress data (Ghosh, 2023). AI can also be used to speed up simulations, simulate complex systems, and improve efficiency (Kung & Lussier, 2023). Machine Learning can address many problems such as "classification, regression, and clustering" and can be used "to produce abstract mathematical conjectures [or] design artificial life forms" (CCA, 2022, p. 10). In the future, AI systems may even be used "to organize research or propose designs", suggesting AI's range of tasks could include "hypothesis generation to interpretation and analysis" or undertake discovery (CCA, 2022, p. xv). In these ways, AI may improve research analysis and determine novel areas of R&D (CCA, 2022; Kung & Lussier, 2023).

In the public sector, the Government of Canada has expressed interest in using AI to assist in or make administrative decisions to enhance the provision of services (Treasury Board, 2023a). AI has the advantage of processing large amounts of data and can complete repetitive tasks faster than humans, allowing humans to dedicate time to higher-risk decisions that require critical analysis, thereby reducing operational costs and netting economic benefits (Alexopoulos et al., 2019; Daly & Orct, 2022; Zuiderwijk et al., 2021; CSPS, 2021b). Algorithms can be used to make predictions, identify patterns, and determine issues as well as solve the 'human errors' of decision making, such as subjectivity and inconsistency in decisions (Daly & Orct, 2022; Loewen 2021b; CSPS, 2021b). In addition, Loewen (2021b) remarks that employing AI in the administrative state is useful precisely because the state is not designed to learn as efficiently as machines. Humans require time to reflect on their decisions and monitor their outcomes to incorporate what they've learnt into their mental models.

RISKS

The widescale adoption of AI also presents risks. The following section explores some of these risks and challenges.

LACK OF TRANSPARENCY: THE BLACK BOX PROBLEM

AI systems today should be thought of as distinct from other technologies because of what many call 'the explainability problem' (CPSC, 2021a; Brookfield Institute, 2018) or 'the black box problem of AI' (Smuha, 2021; Haas, 2020). That is, the inability to see how AI systems process data and make their decisions (Smuha, 2021; Blouin, 2023). Unlike conventional computers that are programmed with a distinct set of rules, ML lets the AI system write its own rules based on large data sets (CSPS, 2021a). Indeed, it was the breakthrough development of computer hardware and algorithms—enabled by increased availability of big data and cheaper production costs—that pulled AI research out of its decades-long stalemate and ushered in the next generation of AI that we know today (CSPS, 2021a; Attard-Frost et al., 2023). Rather than programming the AI with a virtually infinite list of rules (which is an impossible human task), ML feeds the AI system with large sums of data and programs the machine to determine the best rules—or algorithms—to accomplish a specific task.

These developments in ML provide capacity for pattern detection that far surpasses human ability. DL-trained systems can detect patterns and relationships in data that humans are likely to miss, opening up a range of problem-solving applications. All of this, however, gives rise to the explainability or black-box problem: a lack of understanding as to how AI models detect relationships, generate predictions, and make decisions, which creates challenges for interpreting and communicating the results of the algorithmic output. This is due, in part, to the fact that certain types of successful AI models, such as DL or generative AI, are built on complicated algorithms that are not easily understandable (Brookfield Institute, 2018). DL models employ neural networks that are many layers deep with millions of active permutations, and with connections that change at random or heuristically every second (Manheim and Kaplan, 2019). These processes are often not understandable or interpretable by humans, which raises questions about the ethical use of the technology, the accuracy of evaluations, and ability to verify compliance with industry and/or government standards (Smuha, 2021). These same concerns are the impetus behind work to create

‘Explainable AI’, which would be an AI system that can explain its decision-making process (Manheim and Kaplan, 2019). However, many popular AI systems today, particularly those built on DL techniques or generative AI models, still operate on a black-box model (CCA, 2022).

ALGORITHMIC BIAS

AI systems built on ML techniques present the risk of algorithmic bias. As mentioned, ML techniques work by using large data sets for training models. Chat GPT, for instance, was trained on text databases from the internet that amounted to 570 GB, or 300 billion words (Hughes, 2023). When biases are embedded into these large data sets, the machine learns and reproduces those same biases, but in a manner which may seem objective at first glance (Manheim and Kaplan, 2019). For example, in 2015, Amazon sought to automate its recruitment process only to find out that the AI hiring tool disproportionately favoured men because the model was trained to identify patterns based on previous hiring decisions – the majority of which were men (Dastin, 2018). Microsoft’s “teen-talking AI chatbot”, called Tay, began making racist tweets in just a few hours because of its ability to mimic and learn from large amounts of data (Manheim & Kaplan, 2019). In addition, facial recognition technologies have been flagged as unable to recognize dark skin (Brookfield Institute, 2018). Referred to as ‘algorithmic bias’ in the literature, AI, “when let loose on a new set of facts” (Manheim and Kaplan, 2019, p. 158) magnifies embedded biases on a much larger scale.

PRIVACY

AI amplifies previously unresolved privacy issues of Big Data and the Internet of Things. Indeed, Big Data is the enabler behind the advancements in algorithms and ML techniques used today (Risse, 2019). However, there are serious privacy and security concerns with the collection of personal data for use in training datasets or in algorithmic models. (Tessono, 2022). Concerns of the use of private information is amplified when companies retain proprietary ownership of the data (Risse, 2019). Depending on what the technology is used for, private data such as biometric or health data, education, banking, employment, political preferences, and intimate correspondences may be shared and used unknowingly (Manheim and Kaplan, 2019). For example, in 2015, DeepMind Technologies (an Alphabet subsidiary), received 1.6 million patient records from the U.K. National Health Service’s (NHS) Royal Free London NHS Trust through a data-sharing agreement, which the firm declared it would use for a monitoring and diagnostic app for kidney injuries. However, DeepMind Technologies obtained this information without patient consent, and the NHS shared far more information than was initially publicly announced (Brookfield Institute, 2018; Hodson, 2016). According to the findings of the U.K.'s Information Commissioner's Office (ICO) investigation, the NHS's Royal Free Trust failed to comply with data protection principles set out in the U.K.'s Data Protection Act, such as the handling and collection of patient data without informed consent, and the disproportionate and excessive amount of data that was processed relative to the purposes for which the data would be used (Denham, 2017). The findings of the ICO's investigation do not directly implicate DeepMind, as the ICO maintains that the Royal Free Trust's role in this case, was one of a "data controller," and "it is therefore the Royal Free who is required to take the steps…necessary to achieve compliance with the Act" (Denham, 2017, p. 2).

Over-sharing data, intentionally or otherwise, raises an additional risk of unfairly exposing citizens to users who possess the ability to re-identify previously anonymized data (Manheim and Kaplan, 2019). AI's ability to identify patterns from seemingly unrelated sets of data enables it to uncover personally identifiable information by triangulating data points which have been processed or collected by AI systems (Tessono, 2022). For example, technology exists to reverse-identify hospital patients through anonymous billing records and through credit card metadata (Manheim and Kaplan, 2019). Thus, while data first collected by an AI system may indeed be anonymized, sophisticated AI systems are able to re-identify the data using triangulation techniques across vast datasets. This risk, together with the NHS-DeepMind example, calls for rigid provisions to protect privacy and govern the use of anonymous data in order to protect citizens.

THREATS TO CYBERSECURITY

Employing AI can heighten cybersecurity risks by assisting hackers to get through firewall defences (Manheim and Kaplan, 2019). Techniques such as key-stroke logging, intrusion via viruses or worms, or brute-force hacking can be performed more easily with the deployment of AI technologies. The success and capabilities of AI systems are made possible through the massive generation and availability of data. It is possible that hackers can get access to sensitive data if AI systems are not built with the appropriate digital security mechanisms to protect networks and computer systems from malicious attacks. Furthermore, AI systems could be made more vulnerable to attacks if the data used in the design and training of an AI system has been "poisoned" (made inaccurate), or if the AI tool has been trained incorrectly (Canadian Centre for Cyber Security, 2022). While beyond the scope of this paper, further research/reading which examines the cybersecurity-related risks of AI is required for a more comprehensive understanding of these nuances.

DEMOCRACY, ACCOUNTABILITY, AND TRUST

The use of AI in government may create problems for democratic accountability, due process, and the rule of law (Donahoe and Metzger, 2019). These issues have already materialized in several US jurisdictions which apply AI in areas such as parole, sentencing, social services eligibility, and hiring decisions (Dohanoe and Metzger, 2019). The opaque nature and explainability problem of AI technologies create issues for transparency and fairness, especially if the people officially held responsible for particular decisions "do not understand the basis on which the algorithm is making choices" (Donahoe and Meztger, 2019, p. 117). Risks to democratic principles of transparency and accountability are worsened if it is unclear who has control and which human would be ultimately held accountable for decisions (Zuiderwijk et al., 2023; Bullock, 2019). These challenges may contribute to an erosion of public trust in government (Zuiderwijk, 2021) and in democratic institutions (Manheim and Kaplan, 2019). When decision-making processes are not transparent or inclusive, or if they are perceived to be unfair or a violation of privacy, trust in government may be adversely affected (Zuiderwijk, 2021). A decline in trust in government raises concerns about the legitimacy of democracy and the functioning of its institutions. As the Institute on Governance's Action Learning Project points out, trust "is the basic currency for the legitimacy that impacts the government's ability to lead, govern, and or/to respond to a crisis" (Institute on Governance, 2023, p. 11). Thus, the opportunities and risks of using AI in government must be carefully considered as a means to serve and to promote the public good, and to uphold principles of good governance: legitimacy, transparency, accountability, and fairness for all citizens.

The use of AI in government science has implications for both trust in government as well as in the scientific enterprise. Thirteen federal departments and agencies are considered based in science. These departments employ a wide range of scientists, engineers, and technicians who pursue science to advance the mandates of those departments and to inform robust decision-making in the public's interest. Government scientists are also members of a broader scientific community with global reach. Scientific integrity is achieved through the pursuit of knowledge according to a series of commonly held beliefs and the demonstration of certain research practices and protocols. If AI applications used to advance government science are perceived to be ambiguous, unfair, non-consensual, or an invasion of privacy, their use may call into question the legitimacy of that department, in addition to the ethics and integrity of the individual scientists.

These challenges are further augmented by the technological realities of the present: the proliferation of misinformation on social media, the industry practice of passive data collection from our devices and internet traffic, the decline in scientific literacy amongst adult Canadians, and recent anti-intellectualism movements (Bal, 2022). However, scientists also have the opportunity to use AI to promote trust in science and contribute to increased scientific literacy. For example, AI could be used to improve scientific communications by generating plain-language summaries of results (OECD, 2023). This can work to help improve transparency, advance the public's understanding of the scientific process, and, in turn, foster trust in science. At the same time, AI can also be used to verify scientific claims and serve as a 'fact-checker', thanks to its ability to canvas large (and unstructured) datasets at a rate and scale that would be far too laborious and resource intensive to perform manually (Wang, 2023).

PUBLIC PERCEPTIONS OF AI

In 2020, Innovation, Science, and Economic Development Canada (ISED)'s Advisory Council on AI launched The Public Awareness Working Group (the Working Group), with a mandate to investigate public awareness and trust in AI and identify pathways for engagement. The Working Group set out to gather public perceptions of AI and familiarity with AI (AI literacy) through (1) an online, national survey (developed together with Nanos Research), and (2) a series of 19 pan-Canadian workshops (ISED, 2022). The recently published findings of the Working Group's activities revealed the following opinions held by Canadians on AI technology (ISED, 2022) :

The online survey, which collected responses from 1200 Canadians, revealed that:

- Respondents were almost 7 times more likely to believe in the positive impact of AI in Canada, and 4 times more likely to believe that they will experience a positive impact themselves (p. 26).
- Survey respondents are most hopeful about AI applications in manufacturing, transportation, and banking, and most concerned with AI applications in law enforcement and in the workforce (p. 29).
- Respondents cited the following as their primary concerns related to use of AI: job loss, privacy, security, hacking, and malfunctions, in that order.

- 48% of respondents agreed or strongly agreed that AI is developed in an ethical manner and 29% were unsure (p. 22).
- 51% of respondents
- reported familiarity with AI, 21% with some familiarity, and 20% and 8% reporting 'somewhat not familiar' and 'not familiar', respectively (p. 21).

The series of 19 workshops, which involved 437 Canadians, revealed:

- 74% of workshop participants believe in the social benefits of AI (p. 28).
- 71% of participants believe that AI systems can be trusted if governed appropriately by public authorities (p. 32).
- 42% of workshop participants believe the benefit of AI outweighs the costs, 41% had no opinion, and 17% believe the costs outweigh the benefits of AI (p. 27).
- In the areas of healthcare, education, justice, and administrative services, participants identified top ethical issues of bias and discrimination, privacy and data protection, and transparency and explainability as key concerns of AI use (p. 31). Moreover, participants expressed concerns of over-reliance and over-valuation of AI systems capabilities in matters dealing with peoples' lives.
- Participants remarked that AI literacy amongst the public is insufficient, and that AI literacy education must include knowledge of the technology and awareness on engaging with the critical and ethical aspects of AI (p. 25).

These findings indicate that perceptions of AI amongst those consulted are positive. Nevertheless, the responses to the survey and in the workshops depict a general public that expects humans to remain involved in decision support systems concerning humans. However, the number of people who responded to survey questions as 'unsure' reveals a potential gap in AI literacy, and that with increased AI literacy or increased exposure to AI systems, these opinions may trend towards either more positive or more negative views.

The Working Group notes that the findings collected in the survey and in the workshops may not be representative of the general Canadian population. The open criteria for participating in the workshops resulted in a self-selection of participants that had a higher rate (91%) of university-level education than the general Canadian population. Additionally, due to limited resources and scope, targeted population sampling and non-internet-based methods (such as telephone surveys) were excluded, resulting in some populations (such as those most likely to be affected by AI) being under-consulted in the survey and workshops. In addition, the Working Group details that the COVID-19 pandemic introduced further, unique barriers that placed constraints on the generalisability of the report's findings.

The concerns identified in the Working Group's workshops and survey reflect similar concerns held around the globe to mitigate the negative and unintended consequences of this technology without stifling its potential to deliver social benefits (ISED, 2023). More recently, there have been calls from private industry to halt further developments of AI systems until robust governance systems are put in place (Future of Life, 2023). Indeed, Smuha (2021) has argued that amidst international competition to become world-leaders in AI, the 'race to AI' has also brought about "a race to AI regulation."

PART II: GOVERNANCE IN CANADA

At the time of writing, Canada's regulatory framework specific to AI is comprised of the following: (1) the proposed *Artificial Intelligence and Data Act (AIDA)* as part of Bill C-27 the *Digital Charter Implementation Act 2022*, which applies to private sector entities and does not apply to government (Attard-Frost et al., 2023), and (2) The Treasury Board *Directive on Automated Decision-Making (2019)*, a mandatory policy instrument that applies to departments as defined in the Financial Administration Act.[1] The Treasury Board Directive on Automated Decision Making (hereafter known as "The Directive") works in tandem with a set of related guidelines and instruments: The Algorithmic Impact Assessment Tool (AIA), the List of Qualified Artificial Intelligence (AI) Suppliers, section 4.5 of the Guideline on Service and Digital, and the Government of Canada's interdepartmental guiding principles on the "Responsible Use of Artificial Intelligence." The Directive supports the Treasury Boards' broader Policy on Service and Digital, which details a set of guidelines for digital government across Government of Canada organizations (2019).

[1] Some notable exceptions include: the Canada Revenue Agency (CRA), the Office of the Auditor General of Canada, Office of the Chief Electoral Officer Office of the Commissioner of Lobbying of Canada, Office of the Commissioner of Official Languages, Office of the Information Commissioner of Canada, Office of the Privacy Commissioner of Canada, and the Office of the Public Sector Integrity Commissioner of Canada

The following sub-sections discuss the AIDA and the Directive, along with the Directive's related policy instruments, in further detail.

THE ARTIFICIAL INTELLIGENCE AND DATA ACT (AIDA)

In June 2022, the federal government introduced the Artificial Intelligence and Data Act (AIDA) as part of Bill C-27, the Digital Charter Implementation Act, 2022. The AIDA represents the first step towards regulating the development and operation of AI technologies by commercial and individual actors.

As defined in section 39.4 of the AIDA, the purpose of the Act is to "regulate international and interprovincial trade and commerce in AI systems by establishing common requirements, applicable across Canada, for the design, development and use of those systems; and to prohibit certain conduct in relation to AI systems that may result in serious harm to individuals or harm to their interest" (House of Commons, 2021). This scope indicates that it may be left to provinces to implement AI legislation for activities happening exclusively in individual provinces[2].

The Act seeks to address a range of harms and adverse impacts by prescribing specific requirements for firms at each stage of the technology lifecycle. The factors that determine which systems would be considered 'high-impact' include: a) harms to health, safety, human rights based on intended purpose and possible unintended outcomes, and b) severity of harms, scale of use, nature of the harms already committed, degree to which risks are regulated under another law, asymmetry of economic/social situations, and whether or not opting out from the system is possible (ISED, 2023).

In 2023, Innovation, Science, and Economic Development Canada (ISED), released the [Companion document] to the AIDA, which notes important areas of interest to the government with regards to high-impact AI systems: "screening systems impacting access to services or employment, biometric systems used for identification and inference, systems that can influence human behaviour at scale, systems critical to health and safety" (ISED, 2023). Firms that either design, develop, make available for use, or manage the operations of an AI system would be expected to establish their own internal accountability and enforcement mechanisms to ensure compliance with their specific requirements under the AIDA.

[2] IOG was unable to verify this statement with ISED.

The Companion document notes that in the initial years after the AIDA takes effect, the emphasis will be on establishing guidelines, education, and assisting businesses in complying with the Act's obligations. Enforcement mechanisms such as regulatory non-compliance penalties, prosecution of regulatory offences and criminal offences are provided by the AIDA, but with no clarity on when they will come into effect.

The Ministry of Innovation, Science, and Industry (ISI) would administer and enforce the AIDA (with the exception of prosecutable offences), and a newly created AI and Data Commissioner would support the Minister of ISI in their responsibilities, work with regulators in ensuring regulatory capacity, and monitor systemic effects of AI systems to inform policymaking.

The AIDA underwent its second reading in the House of Commons on April 24, 2023, and is now being considered by the Standing Committee on Industry and Technology[3]. It is expected that following Royal Assent, the AIDA will come into effect in 2025, providing for two years for developing regulations that will offer more detail on key items in the Act, such as definitions of systems considered high impact, the specific requirements proportionate to the level of risk, obligations for monitoring, and additional detail on enforcement and compliance tools available to the Minister of ISI.

THE TREASURY BOARD DIRECTIVE ON AUTOMATED DECISION-MAKING

In 2019, the Treasury Board Secretariat of Canada published a *Directive on Automated Decision-Making* ("the Directive") designed to ensure the responsible and ethical use of automated decisions systems by federal government institutions. The Directive "applies to any system, tool, or statistical model used to make an administrative decision or a related assessment about a client" and "only to automated decision systems in production, [excluding] systems operating in test environments" (Treasury Board Directive, s. 5.1, 5.2). "All institutions subject to the Policy on Service and Digital unless excluded by specific acts, regulations, or Orders-in-Council" are subject to the Directive (Treasury Board, 2023a, s. 9.1). Further, in all cases, the Directive only applies to systems introduced or developed after April 1, 2020 (Deshaies, 2021). While not explicitly stated in the Directive, Medeiros and Beatson (2022) interpret that businesses who sell or license automated-

[3] This paper was written in August 2023. IOG was unable to validate the content of this paragraph with ISED.

decision system technologies to the federal government are subject to the requirements set out in the Directive.

The Directive sets a number of requirements that assistant deputy ministers are subject to prior to deploying any AI technologies (Treasury Board Directive, 2023, s. 6). First, organizations are required to complete the Algorithmic Impact Assessment Tool, which is a mandatory risk assessment questionnaire of roughly 80 questions that is designed the identify the risk level of an automated decision-making system on "the rights (including equality, dignity, privacy, and autonomy) and health or well-being of individuals/communities, the economic interests of individuals or communities, and the ongoing sustainability of an ecosystem" (Government of Canada, 2023a). AI technologies are classified into one of four impact levels (ranging from low to high impact), which each level obligated to undertake a unique and additional series of analyses designed to protect the rights listed above. Results of the assessments are available via the Open Government data portal.

Second, organizations employing AI tools must adhere to a set of transparency standards that include a) provide notice (prominently and in plain language) to relevant stakeholders/clients of the use of such automated decision systems, b) provide meaningful explanations alongside automated decisions to the affected individual, c) provide the Government of Canada with access to components, d) release source code owned by the Government of Canada, and e) document automated decisions in accordance with the Treasury Board Directive on Service Digital (Treasury Board, 2023a).

Third, the Directive outlines Quality Assurance requirements, such as: required testing for bias, ongoing monitoring to mitigate unintentional outcomes, validation of data quality, peer review, legal consultations, Gender-based Analysis Plus (GBA+), and measures to ensure data is traceable, protected, and lawfully used. Other quality assurance requirements include: employee training in the design, function, and implementation of the system, risk assessments during development of the automated system, strategies to support IT and business continuity management, and ensuring human intervention. Finally, the Directive provides recourse for clients to challenge the administrative decision (s. 6.4.1) and requires updated reports on the effectiveness and efficiency of the decision system in meeting program requirements (s. 6.5.1). Consequences of non-compliance are listed in the Framework for the Management of Compliance and appendaged in the Directive (Treasury Board, 2023a).

To support compliance with requirements laid out in the Directive, the Treasury Board together with Public Services and Procurement Canada (PSPC) released a list of pre-approved suppliers of "responsible and effective AI services", which federal departments and agencies can use to streamline the process of procuring AI systems (Treasury Board, 2023b). In addition, section 4.5 of the Guideline on Service and Digital provides additional instructions which relate to the responsibilities of departments (s. 4.5.1 – s. 4.5.2) and details additional considerations for procuring and implementing automated systems (s. 4.5.3). Finally, the Directive is supported (and informed by) the Government of Canada's guiding principles on AI (Government of Canada, 2023c), which were developed and adopted by Canada and other nations at the Digital 9 Summit in 2018 (Coulson, 2023). They are:

1. understand and measure the impacts of using AI;
2. transparency in how and why AI is being used;
3. meaningful explanations of AI decision-making with opportunities for recourse;
4. openness in the sharing of data and source code while protecting personal and confidential information;
5. ensure sufficient training for government employees in developing and using AI.

PART III: AI IN FEDERAL SBDAS

To collect more information than what is available in the available literature and to further inform this paper of the uses and applications of AI in federal SBDAs, a short questionnaire was presented to members of the federal research and scientific community. The survey collected information from 49 respondents on current or potential uses and applications of AI in respective organizations, including whether opportunities exist for knowledge sharing and upskilling, what evaluative and accountability measures are in place, whether AI has a place in scientific mandates, and how the ethical and practical considerations of using AI are being responsibly addressed. A summary of the results is presented below.

Eighty-six percent of respondents report that their organizations are currently using or will be using AI. Of that 86%, most uses (54%) will or are contributing to R&D functions and a smaller subset (roughly 19%) are contributing to administrative functions. Twenty-seven percent of respondents report AI is used for 'other' functions, such as compiling information for analysis, writing of reports and project reporting, scientific monitoring, reviewing scientific literature, scientific searches, job applications, natural language processing of text survey data, and policy development.

Other examples of AI use include intelligence document processing (IDP), reading structured text, performing scientific searches or scouring historical data, data interpretation and analysis of large datasets, summarizing and translating documents, creating code and excel formulas for data analysis, identification of duplicate and near duplicate documents in ATIP operations, supporting statistical or process-based models, garnering opinions, as a supplementary method for searching for programming resources, explaining concepts related to statistical methodologies in lay terms, long-form correspondence, writing government documents, classifying and processing data, and modelling and making predictions. These use-cases of AI in government science are consistent with the wider literature describing AI technology for scientific applications.

Many respondents (64%) were unsure or unaware of policies or programs being developed within their respective departments or agencies to guide the use of AI. Twenty-two percent of respondents note within their departments, there are working groups and/or individuals involved in developing such policies, such as information technology units or in the case of CFIA, in the AI lab. Two respondents mentioned that they rely on the Treasury Board's policies and directives. One respondent says they consult the Public Service Commission of Canada for guidance on AI use, while two others note that these activities are coordinated/centralized within the Office of the Chief Data Officer.

Roughly twenty-seven percent of respondents explain that they are aware of some courses, workshops, or guidelines on AI in their department or agency. The responses indicate that some of these trainings have not been delivered in formal programs, but on a project-specific or unit-specific basis that train scientists on how to use AI tools for their distinct research needs. In one case, scientists were provided with safe experimentation guidelines. One respondent specifies that they have received webinars explaining the uses, limitations, and appropriability of ChatGPT, while another mentions the courses provided by the Canada School of Public Service (CSPS) on AI. Over half (64%) of respondents report that training should be available using AI systems in their respective department or agency. The majority explained that training for scientists or researchers in using AI systems is largely done on an ad-hoc or independent study basis. Those scientists and researchers working with AI tools are largely self-trained or have learned on the job, or in the case of one participant, have collaborated with scientists in other countries to advance their work.

In general, organization-wide strategies on AI use for scientific activities remain in preliminary stages. Of 47 respondents, 23% percent report their department or agency has plans to integrate AI into overall research and scientific objectives, but most of these plans are still exploratory in nature and have yet to move beyond initial discussion stages. Nineteen percent of respondents report that their department or agency has plans for a long-term AI strategy.

Seventy percent of respondents note that their organizations are not formally evaluating, assessing, and reporting on the use of AI, or are not aware of any such activities occurring. Twenty-two percent report having evaluative and reporting mechanisms, which range from simple reviews of the response, comparison with manual methods, or survey feedback forms to validation tests and specific performance metrics of AI activities unique to the research activity. Two respondents explain they are evaluating and reporting on the use of AI according to time saved by automation of tasks, effectiveness, ease of use, and cost of training. The responses indicate that evaluative mechanisms exist largely on a unit-specific or project-specific basis. Agency or department-wide guidelines still need to be created but are expected to be forthcoming as familiarity with AI technology and its functionality grows.

With respect to considerations of the risks of privacy, bias, accountability, and transparency, respondents report a variety of approaches to addressing these concerns. Three respondents report that their departments lean on the Treasury Board's assessment tool, while others communicated that these concerns are addressed via internal ethics courses or talks, informal discussions with peers, assessed on a case-by-case basis, or rely on existing best practices relating to transparency and privacy. Many of the uses or applications of AI do not deal directly with human data, and therefore do not possess the same ethical implications as other uses. Overall, the responses detail that federal SBDAs are still in the early stages of developing the policy mechanisms for mitigating these risks. As the use of AI is considered further, SBDAs are likely to implement more detailed frameworks to facilitate ethical practices in the near future.

The responses to the survey provide a glimpse into the use of AI in federal SBDAs, but they do not describe the whole picture. In general, the survey responses illustrate a federal scientific community that is still in the early investigative stages of understanding the appropriate role of AI technologies in their R&D or RSA activities. A more representative study with a larger number of participants is necessary to build a better picture of the breadth of AI uses and applications in government science. In addition, this survey did not ask respondents to identify their department or agency to ensure anonymity. As such, no data was collected on the distribution of responses according to department. A future study could filter or factor for this data component to understand the range of AI activity across all 13 science-based departments and agencies.

PART IV: DISCUSSION

Both the AIDA and the Directive demonstrate the Government of Canada's intention to mitigate the risks of AI while still recognising that AI has the potential to serve broad social benefits. Many of the challenges identified earlier in the paper – such as issues of transparency, bias, and privacy – are addressed in both documents. Since 2021, 11 algorithmic impact assessments have been published on the Open Government data portal on behalf of Veteran Affairs Canada, Public Health Agency of Canada (PHAC), Immigration, Refugees, and Citizenship Canada (IRCC), and Economic and Social Development Canada (ESDC) (Government of Canada, 2023d). The impact assessment gives each automated decisions system a rating of 1 to 4: A rating of 1 indicates that the automated decision will likely have *little to no impact* on the rights (including equality, dignity, privacy, and autonomy), health and wellbeing, and economic interests of individuals or communities or the sustainability of an ecosystem, and if there are impacts, these impacts are brief and reversible (Treasury Board, 2023a). A rating of 4 indicates that the automated decision will likely have *very high impacts* on the rights, health, well-being, and economic interests of individuals or communities. The impacts of level 4 decisions are often permanent and irreversible (Treasury Board, 2023a). Ratings of 2 and above require the Assistant Deputy Minister or any other person named by the deputy head responsible for the AI system to: have the automated decision system peer-reviewed by qualified experts, complete a Gender-based Analysis Plus (GBA+), provide notice in plain language (with more specific requirements for Level 3 and 4 systems), a meaningful explanation of the role of the automated system (including the data and criteria used), and documentation on the design and functionality of the system for training purposes (with required training courses for Level 3 and 4 systems) (Treasury Board, 2023a).

Level 3 and 4 systems require humans to remain involved in the decision-making process and the final decisions to be rendered by a human. In addition, level 3 systems require approval from the Deputy Head and level 4 systems from the Treasury Board, while level 1 and 2 systems require the same approvals as other IT systems. To date, all 10 projects assessed have received a score of 2 to indicate the automated decision will likely have *moderate impacts* on the rights, health and wellbeing, and economic interests of individuals or communities and that the impacts of the automated decision will be likely short-term and reversible.

GAPS

The Directive and its related guidelines serve as the primary instrument governing the use of AI by federal government. While the AIDA focuses on private sector entities, its provisions may still remain relevant to federal employees (to be determined). As mentioned above, businesses who license or sell automated-decision system technologies must meet the requirements in the Directive. If an AI solution or technology is procured by departments, departments remain responsible for ensuring the system meets the requirements of the Directive.

Is there a policy gap that exists when it comes to businesses that design and develop an automated system for the public sector? For example, Clearview AI (an American facial recognition software company) and the Royal Canadian Mounted Police (RCMP) became the subject of a 2021 investigation by the Privacy Commissioner of Canada (OPC) because of the company's unlawful collection, use, and disclosure of personal information (OPC, 2021). Clearview was found in violation of Canadian privacy laws when it created a database of more than three billion images scraped from the internet without consent. The RCMP was found in violation of the Privacy Act when it collected the unlawful personal information from Clearview AI, and because it failed to ensure it was using a lawful database (OPC, 2021). The RCMP had argued that under current federal privacy laws, it was under no explicit obligation to corroborate the legality of the methods of data collection by private sector partners (OPC, 2021). Alongside their report, the OPC called on Parliament to amend the Privacy Act accordingly to reflect this obligation.

To avoid situations like the Clearview AI example, the Directive sets out requirements for the collection and governance of data. Sections 6.3.3 requires that data collected and used for the automated decision system remains in accordance with the Policy on Service and Digital and the Privacy Act. Section 6.3.4 provide for measures that make sure the data generated and used by the automated system are "traceable, protected and accessed appropriately, and lawfully collected, used, retained, and disposed of" (Treasury Board, 2023a). This gives the department the authority to collect the data. The OPC's findings in the Clearview investigation indicate that government institutions have additional obligations to ensure that the AI systems they use were developed using lawful data collection techniques and with appropriate consent, whether those systems were procured from third parties or developed internally.

The Directive's scope indicates that the Directive applies to the use of "any system, tool, or statistical model" (Treasury Board, 2023a, s. 5.1). There are also many requirements in the Directive which apply prior to production (prior to use) of an automated decision system, such as conducting an algorithmic assessment or quality assurance requirements. This indicates that the design and development of automated systems for use in federal departments is meaningfully addressed in the Directive. The companion document to the AIDA notes that the 'design' of an AI system includes "determining AI system objectives and data needs, methodologies, or models based on those objectives" (ISED, 2023). The 'development' of an AI system "includes processing data sets, training systems using the datasets, modifying parameters of the system, developing and modifying methodologies or models used in the system, or testing the system" (ISED, 2023).

The majority of automated decision systems used in federal departments have been developed internally by departments. However, many of the general AI systems on the market today have largely been driven by commercial incentives and market demand in Big Tech companies (CSPS, 2021a). Designing and developing an AI system for commercial use will look different from designing an AI system for public sector use because of the fundamentally different incentives underpinning each. The public sector faces different constraints and is characteristically concerned with public benefit, while the private sector operates largely on economic incentives. As Loewen observes, "we need more AI being built to the specs of the public sector, not just the private sector" (CPSC, 2021a).

The particular example of Clearview points to an additional obfuscation; a federal organization using a software licensed by a non-Canadian company. The Directive does not clarify rules around using AI systems developed by non-Canadian firms, who are presumed to be subject to regulations and requirements in their respective jurisdictions. In such cases, risks to privacy and confidentiality are amplified because the personal information of Canadians is being stored in servers in another jurisdiction outside of Canada. Because many of the details of the AIDA will be sketched out in regulations, it is still being determined whether these concerns will be addressed and how they intersect with existing Canadian laws that deal with foreign enterprises conducting business in Canada.

On a more granular level, one of the concerns raised with previous versions of the Directive was that it only applied to services provided externally of government (Scassa, 2022). This would mean internal uses, such as for hiring decisions and performance reviews, are excluded.

The Directive has since been updated (April 25, 2023) with this provision removed, but it does not explicitly clarify this adjustment in its scope. Instead, the Directive relies on the use of the term 'client' to convey the applicable scope, with a definition of the term 'client' provided for in the Policy on Service and Digital.

AI SYSTEMS APPLICATIONS IN SCIENCE

What about AI systems applications in science? The Directive regulates the use of AI systems only when these systems are used to make administrative decision about a client (Bitar et al., 2022). Administrative decisions are defined as "any decision that is made by an authorized official of an institution as identified in section 9 of this directive... that affects legal rights, privileges or interests" (Treasury Board, 2023a). This indicates that the use of AI systems for non-administrative decisions—decisions which do not impact the legal rights, privileges, or interests of another person—are exempt from the requirements in the Directive. Some applications of AI in science, such as for design and discovery purposes (CCA, 2022) or for scientific writing (Salvagno et al., 2023) may fall beyond the scope of the Directive. In these cases, it would be left to individual departments or agencies to determine the rules for using AI systems for scientific applications or for research purposes. In fact, a few individual departments have released their own guiding principles relating to specific aspects of the department's range of activities:

- Health Canada's "*Good Machine Learning Practice for Medical Device Development*" (2021),
- Statistics Canada's "Framework for Responsible Machine Learning Processes" (2021), and,
- Innovation, Science, and Economic Development Canada's (ISED) "Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things" (ISED, 2021).

These activity or industry-specific guidelines share some of the same commitments on responsible AI use: a) robust assurances for data quality, data management, and reference standards for datasets, b) specific guidelines on the appropriate explainability level of the model, and c) monitoring and assessment protocols to ensure models meet department-specific requirements. Collectively, these guidelines demonstrate initial efforts to ensure that the obligations for AI are context-specific to the department or agency in question.

Based on the responses to the survey, to-date uses of AI in federal SBDAs remain exploratory in nature, and only a few deal directly with human data and personal information. Moreover, the majority of AI uses relate specifically to R&D functions with less than a third contributing to administrative capacities. Many use-cases of AI in R&D are also performed informally, on an ad-hoc basis. To this end, there are a number of individuals and pockets of working groups involved in developing policies and programs to help guide the use of AI, in addition to the Treasury Board's guidelines. This suggests that there is an appetite for establishing guidelines and strategies for the use of AI in department-specific R&D applications, but more information is needed to ascertain the functionalities and limitations of AI systems in scientific applications. Taken altogether, the knowledge infrastructure for facilitating and guiding the use of AI in most SBDAs has yet to be established but is expected to concretize clearer guidelines and a better understanding of the intended applications of AI in the coming years.

When considering the use of AI, the federal research and scientific community should be aware of a potential trade-off between using model-driven techniques and simpler forms of AI systems. Model driven techniques (such as DL and neural networks) have the benefit of predictive accuracy, but also suffer from the 'black box' problem, while simpler forms of AI systems (such as statistical ML) retain explainability naturally but are not as accurate (CCA, 2022). Particularly because the scope of the Directive may be extraneous to certain applications of R&D and RSA, explanations of the automated decisions may not be required. Therefore, institutions conducting R&D or RSA in areas that fall beyond the scope of the Directive may need to develop institution-specific requirements guiding the use of AI in the scientific process, including the degree to which 'black-box' systems are permissible and in which contexts. The Guideline on Service and Digital provides further guidance under section 4.5.3 when selecting AI models "if machine learning is used in the automation of decision-making" (Government of Canada, 2023b, s.4.5.3), which SBDAs may consult. Section 4.5.3 also provides for greater clarity on the nuances of using certain types of AI models over others, such as the benefits and disadvantages of using DL and neural networks over simpler algorithms, and on considerations of interpretability and explainability.

EXISTING LEGAL AND POLICY FRAMEWORKS

Beyond the policy instruments discussed here, Canada already has legal frameworks that could apply to the many uses of AI. For example, the Personal Information Protection and Electronic Documents Act (PIPEDA) and the Privacy Act regulate the use of personal information [4], while other

[4] Until Bill C-27 takes effect, which in addition to introducing the AIDA, seeks to enact *the Consumer Privacy Protection Act and Data Protection Tribunal Act* to modernize and replace the existing PIPEDA.

statutes such as the Canada Consumer Product Safety Act (CCPSA), Food and Drugs Act, and the Motor Vehicle Safety Act (MVSA) enable the development of safety regulations to protect consumers (ISED, 2023). While these frameworks do not explicitly address AI systems, some aspects may be applicable to the use of AI technology.

The collection, use, and disclosure of personal information by federal government institutions is governed through the Privacy Act, and through the Personal Information Protection and Electronic Documents Act (PIPEDA) for private sector organizations carrying out commercial activities. Federal programs engaged in AI system design or development or who make use of AI systems, particularly where personal data is collected and used for training or for carrying out specific tasks, will be subject to the provisions of the Privacy Act. Contrarily, private enterprises that carry out these same tasks must comply with requirements under PIPEDA.

The Canada Consumer Product Safety Act (CCPSA) defines a consumer product as “a product, including its components, parts or accessories, that may reasonably be expected to be obtained by an individual to be used for non-commercial purposes, including for domestic, recreational, and sports purposes, and includes its packaging” (Canada Consumer Product Safety Act S.C 2010, c. 21). If an AI system is used in a consumer product, it’s components may be subject to the regulations set out under the CCPSA.

The Canada Food and Drugs Act (CFDA) attempts to ensure safe food, drugs, cosmetics, medical devices, and therapeutic devices are sold to Canadians and that their ingredients are disclosed and not misleading. The CFDA may be applicable to the use of AI in healthcare. Integrating AI systems into medical devices, for example, may be subject to the regulations set out by Health Canada under the CFDA. With respects to the inspection activities outlined in the CFDA, it may be the case that these activities are supplemented with AI tools, but the Act does not detail specific requirements on the use of AI for inspection purposes.

The Motor Vehicle Safety Act (MVSA) empowers Transport Canada to develop and enforce safety regulations for regulated classes of motor vehicles and vehicle equipment. AI incorporated into automated driving features, control systems, or advanced assistance systems may be subject to the safety regulations and oversight authorities established by the MVSA.

Lastly, scientific integrity policies adopted by federal departments and agencies may be another place where select aspects could apply to the use of AI; they could become a point of departure for

SBDA's looking to develop policies guiding the ethical development and use of AI systems in scientific research. Departments could modify and or expand existing SIP's (OCSA, 2021) to include considerations of AI, or develop a new policy built around the core principles of their departmental scientific integrity policy. Particularly, those aspects of the scientific integrity policy which are concerned with the responsible conduct of research and with monitoring and evaluation may be most applicable to a consideration of AI systems.

Because existing legislative or policy frameworks do not make explicit reference to AI systems, further clarification is needed to discern their applicability. While beyond the scope of this paper, these existing legislative frameworks raise questions regarding the relationship between existing statutes and the policy instruments specific to AI.

DISCUSSION QUESTIONS

1. The drive to elevate innovation across the Government of Canada will increasingly seek to employ AI in projects. Those hoping to use or develop AI within the Government of Canada will increasingly need to develop the general literacy and skill sets required to use or develop AI effectively and responsibly, with a specific focus on the unique risks of AI. What are the key concepts, skills, competencies, or capabilities which are necessary to equip government scientists to use AI systems ethically and responsibly?

2. AI and the internet of things (IOT) are destined to be combined for the benefits of AI to be maximized: what are the unique risks that exist at the interface of AI and IOT that government scientists will need to navigate?

3. AI is an example of a technology that will require an integration of technical knowledge with policy-specific knowledge in order to minimize risks while maximizing scientific potential. How can scientists working at the science-policy interface ensure the effective transfer of knowledge to policy-makers to support evidence-based decision making?

4. How should the calls from private industry and academia to halt further development of AI systems until robust governance systems are in place influence or inform how government should proceed in the use, development, and governance of AI in government?

5. Is there a tension between attempts to regulate AI in the public sector, while simultaneously promoting its uptake and use within Government of Canada?

GLOSSARY OF TERMS

Artificial Intelligence (AI): "Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours or solving problems" (Treasury Board, 2023a).

Automated Decision-System: "any technology that either assists or replaces the judgment of human decision-makers through rules-based system, regression analysis, predictive analytics, machine learning, deep learning, a neural network or other technique" (Treasury Board, 2023a).

Artificial General Intelligence (AGI): an intelligence system that is capable of carrying out all human-level intellectual tasks, which includes learning, problem solving, reasoning, task completion, and is able to switch between a variety of tasks simultaneously (Brookfield Institute, 2018; Allen and West, 2020).

Artificial Super-intelligence (ASI): an intelligence that would surpass and outperform humans in nearly all areas, which goes beyond reasoning and problem-solving to include social skills, general wisdom, and scientific creativity (Brookfield Institute, 2018).

Big data: Very large data sets that can involve billions of records and which require powerful computer-processing to analyze (Allen and West, 2020).

Deep Learning (DL): a form of statistical machine learning that "uses learning algorithms called artificial networks that are loosely inspired by the structure of the human brain. Artificial neurons are connected to one another in layers that rewire and edit themselves on the fly. This approach "allows DL to find patterns in unstructured data, from which it models knowledge representation in manner that resembles reasoning" (Manheim and Kaplan, 2019).

Generative AI: a type of AI that "generates new content [in the form of text, image, audio, or software code] by modelling features of data from large datasets that were fed into the model" (Canadian Centre for Cyber Security, 2023).

Large language models (LLMs) are one class of generative AI that have experienced significant advancements. LLM's involve "developing algorithms and models that can process, analyze, and generate natural language text or speech trained on vast amounts of data" (OCED, 2023). OpenAI's Chat GPT or Google's LaMDA are examples of LLMs which have generated an explosive interest in AI technologies and their applications (Canadian Centre for Cyber Security, 2023).

Machine Learning (ML): a technique "that enables computer systems to learn and make predictions based on historical data…powered by a machine learning algorithm…that is able to improve its performance over time by training itself using methods of data analysis and analytical modelling" (Brookfield Institute, 2018, p. 4).

Model-driven ML: An approach which generates mechanistic models from the data consistent with the data themselves that can be tested against newly generated data" (King & Zenil, 2023, p. 181). The models are mechanistic in the sense that "they can be followed state by state, as in a dynamic system, through a chain of cause and effect" (p. 181).

Narrow AI: sometimes called 'Weak AI', refers to AI that is capable of performing individual or a narrow set of domain-specific tasks. Narrow AI can only do what it is designed to do (Brookfield Institute, 2018).

Natural Language Processing (NLP): "a functionality that enables machines to process, understand, and/or generate audio and textual speech" (Brookfield Institute, 2018, p. 6). NLP is the machinery behind applications such as language translation, chatbots, and AI assistants (Brookfield Institute, 2018).

Neural Networks: "a processing device (algorithms or actual hardware) modelled after the neuronal structure" of the human brain (Brookfield Institute, 2018). Neural networks "learn in layers and build complex concepts out of simpler ones. They break up tasks, identify objects, and apply that knowledge to other activities…which allows computers to learn. Deep Learning and Machine Learning operate through neural networks" (Allen and West, 2020).

Predictive Analytics: "the use of data analytics and machine learning to extract information and learn patterns from data in order to uncover past, present, and future events" (Brookfield Institute, 2018).

Reinforcement learning: a method of training algorithms using "rewards and punishments in the form of functions" (Brookfield Institute, 2018).

Statistical ML: the most dominant and successful form of ML. It is based on complex pattern learning and powerful statistical computation. Statistical ML works by finding regularities in datasets, which "can then be interpreted or studied further" (King & Zenil, 2023, p. 181).

Semi-supervised learning: a method of training algorithms with a combination of labelled and unlabeled data (Brookfield Institute, 2018).

Supervised learning: a method of "teaching a machine learning algorithm by providing a labelled training data set, determining what input features will correspond with learned functions, and providing an example of correct outputs" (Brookfield Institute, 2018).

Training data: data which is used to train ML and DL algorithms. The data can be structured (presented in a standardized format), unstructured (presented without pre-definitive models and contains various types of data ranging from qualitative to quantitative), or semi-structured (presented as unstructured data but which contains tags or labels to enforce order) (Brookfield Institute, 2018).

Unsupervised learning: involves the use of unlabelled training data to which the algorithm must "structure data, discover patterns, classify inputs, learn functions, and produce outputs without external validation or support" (Brookfield Institute, 2018).

BIBLIOGRAPHY

Allen, J. and West D. 2020. The Brookings glossary of AI and emerging technologies. *Brookings Institute.* [Online]. [Accessed August 27 2023]. Available from: https://www.brookings.edu/articles/the-brookings-glossary-of-ai-and-emerging-technologies/

Alexopoulos, C., Lachana, Z., Androutsopoulou, A., Diamantopoulou, V., Charalabidis, Y., & Loutsaris, M. A. 2019. How machine learning is changing egovernment. In*: Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance*, 3-5 April, Melbourne, Australia, pp. 354- 364. Doi: https://doi.org/10.1145/3326365.3326412

Attard-Frost, B., Brandusescu, A. and Lyons, K. 2023. The Governance of Artificial Intelligence in Canada: Findings and Opportunities from a Review of 84 AI Governance Initiatives. Available at SSRN: https://ssrn.com/abstract=4414212.

Bal, R. 2022. Trust, Integrity, and Science Ethics. Government Science and Innovation in the New Normal Discussion Paper. *Institute on Governance.* Available from: https://iog.ca/wp-content/uploads/2022/11/2022-11-04_GSINN_TISE_Discussion_Paper-1.pdf

Bitar, O., Deshaies, B., and Hall, D. 2022. 3rd Review of the Treasury Board Directive on Automated Decision-Making. Available at SSRN: https://ssrn.com/abstract=4087546.

Blouin, L. 2023. AI's mysterious 'black box' problem, explained. *University of Michigan- Dearborn News*. March 6. Available from: https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

Brookfield Institute. 2018. Intro *to AI for Policymakers: Understanding the Shift.* Available from: https://brookfieldinstitute.ca/intro-to-ai-for-policymakers/.

Bullock, J. B. 2019. Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7), pp. 751–761. doi: https://doi.org/10.1177/0275074019856123.

Callaway, E. 2020. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* 588 (7837), pp. 203-204. doi: https://doi.org/10.1038/d41586-020-03348-4.

Canada School of Public Service (CSPS). 2021a. *What is AI? Artificial intelligence is Here Series.* [Online]. [Accessed June 15 2023]. Available from https://www.csps-efpc.gc.ca/events/artificial-intelligence-here-series/index-eng.aspx.

Canadian Centre for Cyber Security. 2023. Generative artificial intelligence. Government of Canada. [Online]. [Accessed August 27 2023]. Available from: https://www.cyber.gc.ca/en/guidance/generative-artificial-intelligence-ai-itsap00041

Canadian Centre for Cyber Security. 2022. Artificial Intelligence. Government of Canada. [Online]. [Accessed August 27 2023]. Available from: https://www.cyber.gc.ca/en/guidance/artificial-intelligence-itsap00040

Canada School of Public Service (CSPS). 2021b. *Citizen Consent and the Use of AI in Government*. *Artificial intelligence is Here Series.* [Online]. [Accessed June 15 2023]. Available from: https://www.csps-efpc.gc.ca/events/artificial-intelligence-here-series/index-eng.aspx.

Council of Canadian Academies (CCA). 2022. *Leaps and Boundaries: The Expert Panel on Artificial Intelligence for Science and Engineering.* Council of Canadian Academies. Available from: https://www.cca-reports.ca/reports/ai-for-science-and-engineering/.

Chow, A. and Perrigo, B. 2023. The AI Arms Race is Changing Everything. *Time Magazine.* [Online]. 17 February. [Accessed June 6 2023]. Available from: https://time.com/6255952/ai-impact-chatgpt-microsoft-google/.

Coulson, S. 2023. How artificial intelligence will change administrative law: The Government of Canada's Directive on automated decision-making. *DLA Piper.* [Online]. [Accessed June 23 2023]. https://www.dlapiper.com/en-ca/insights/publications/2023/05/how-artificial-intelligence-will-change-administrative-law-in-canada.

Daly, P., and Orct, B. 2022. Artificial Intelligence Accountability of Public Administration in Canada: Country Report for the 2022 General Congress of the International Academy of Comparative Law. CanLIIDocs 4250. Available from: https://www.canlii.org/en/commentary/doc/2022CanLIIDocs4250.

Dastin, J. 2018. Amazon ditches AI recruiting tool that didn't like women. *Global News.* [Online]. 10 October. [Accessed June 3 2023]. Available from: https://globalnews.ca/news/4532172/amazon-jobs-ai-bias/.

Denham, E. 2017. RFA0627721 – provision of patient data to DeepMind. *Information Commissioner's Office*. [Online]. United Kingdom. Available from: https://ico.org.uk/media/action-weve-taken/undertakings/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

Deshaies, B. 2021. Responsible use of automated decision systems in the federal government. *Treasury Board of Canada Secretariat.* Available from: https://www.statcan.gc.ca/en/data-science/network/automated-systems

Donahoe, E., & Metzger, M. 2019. Artificial Intelligence and Human Rights. *Journal of Democracy*, 30(2), pp. 115–126. Available from: https://doi.org/10.1353/jod.2019.0029

Faculty of Computer Science. 2023. Artificial Intelligence. *University of Toronto.* Available from: https://web.cs.toronto.edu/research/artificial-intelligence.

Francois, P. 2023*. Geotechnomics: Canadian Perspectives*. [Webinar]. [Online]. Futures Week 2023, Policy Horizons Canada. 17 May 2023.

Future of Life. 2023. Pause Giant AI Experiments: An Open Letter. [Online]. 22 March. [Accessed 3 June 2023]. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Ghosh, A. 2023. How can artificial intelligence help scientists? A (non-exhaustive) overview. In *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*. Organization for Economic Cooperation and Development (OECD). [Online]. OECD Publishing, pp. 103-113. Available from: https://doi.org/10.1787/a8d820bd-en.

Government of Canada. 2023a. Algorithmic Impact Assessment Tool. [Online]. Available from: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.

Government of Canada. 2023b. Guideline on Service and Digital. [Online]. [Accessed June 20 2023]. Available from: https://www.canada.ca/en/government/system/digital-government/guideline-service-digital.html.

Government of Canada. 2023c. Responsible use of Artificial Intelligence (AI). [Online]. [Accessed June 20 2023]. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1.

Government of Canada. 2023d. Open Government Portal. Algorithmic Impact Assessment. [Online]. [Accessed June 20 2023]. Available from: https://search.open.canada.ca/opendata/?collection=aia&page=1&sort=date_modified+desc

Government of Canada. 2019. Policy on Service and Digital [Online]. Available from: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32603

Haas, J. 2020. Freedom of the media and artificial intelligence. Global Conference for Media Freedom, 16 November. [Online]. Government of Canada. [Accessed 3 June 2023]. Available from: https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/human_rights-droits_homme/policy-orientation-ai-ia.aspx?lang=eng

Haenlein, M. and Kaplan, A. 2019. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review.* 61 (4), pp. 5-14. Doi: https://doi.org/10.1177/0008125619864925

Health Canada. 2021. Good Machine Learning Practice for Medical Device Development: Guiding Principles. [Online]. 27 October. [Accessed 20 June 2023]. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/good-machine-learning-practice-medical-device-development.html

Hodson, H. 2016. Revealed: Google AI has access to huge haul of NHS patient data. *New Scientist.* [Online]. 29 April. [Accessed 19 June 2023]. Available from: https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/

House of Commons Canada. 2021. Bill C -27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts* (2022), 1st session, 44th

Parliament (second reading 24 April 2023). Available from: https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

Hughes, A. 2023. Everything you need to know about OpenAI's GPT-4 tool. *BBC Science Focus*. [Online]. 20 June. [Accessed 23 June 2023]. Available from: https://www.sciencefocus.com/future-technology/gpt-3/

Innovation, Science, and Economic Development Canada (ISED). 2023. The Artificial Intelligence and Data Act (AIDA) – Companion document. Government of Canada. Available from: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

Innovation, Science, and Economic Development Canada (ISED). 2022. *Learning Together for Responsible Artificial Intelligence: Report of the Public Awareness Working Group.* Available from: https://ised-isde.canada.ca/site/advisory-council-artificial-intelligence/en/public-awareness-working-group/learning-together-responsible-artificial-intelligence

Innovation, Science, and Economic Development Canada (ISED). 2021. *A Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things.* Government of Canada. Available from: https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/copyright-policy/consultation-modern-copyright-framework-artificial-intelligence-and-internet-things-0

Institute on Governance (IOG). 2023. Trust in Government: Action Learning Project. [Online]. Available from: https://iog.ca/wp-content/uploads/2023/05/Trust-Paper-Final_CW_May-26-2023.pdf

King, R. And Zenil, H. 2023. Artificial Intelligence in Scientific discovery: challenges and opportunities. In *Artificial Intelligence in Science: Challenges, Opportunities, and the Future of Research*. [Online]. OECD Publishing, pp. 181- 187. Available from: https://doi.org/10.1787/a8d820bd-en.

KMPG. 2023. More than a third of Canadian experimenting with Chat GPT, KMPG Canada survey finds. Available from: https://kpmg.com/ca/en/home/media/press-releases/2023/04/us-outpacing-canada-in-business-adoption-of-ai.html

Kung, J. and Lussier, K. 2023. *CIFAR Symposium on AI for Energy and the Environment, Summary Report.* Canadian Institute for Advanced Research (CIFAR). Available from: https://cifar.ca/cifarnews/2023/06/02/report-cifar-symposium-on-ai-for-energy-and-the-environment/

Kuziemski, M., & Misuraca, G. 2020. AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6) 101976. Doi: https://doi.org/10.1016/j.telpol.2020.101976

Manheim, K., & Kaplan, L. 2019. Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law and Technology*. *21*, pp. 106–188. Available at SSRN: https://ssrn.com/abstract=3273016.

Medeiros, M., and Beatson, J. 2022. Bill C-27: Canada's first artificial intelligence legislation has arrived. *Norton Rose Fulbright.* Available from: https://www.nortonrosefulbright.com/en/knowledge/publications/55b9a0bd/bill-c-27-canadas-first-artificial-intelligence-legislation-has-arrived.

Munk Debates. 2023. Be it resolved, AI research and development poses an existential threat. June 23, Roy Thomspon Hall, Toronto.

OECD. 2023. AI language models: Technological, socio-economic, and policy considerations. OECD Digital Economy Papers no. 352. [Online]. OECD Publishing. Available from: https://doi.org/10.1787/13d38f92-en.

Office of the Chief Science Advisor (OCSA). 2021. Model Policy on Scientific Integrity. [Online]. [Accessed August 27 2023]. Available from: https://science.gc.ca/site/science/en/office-chief-science-advisor/model-policy-scientific-integrity

Office of the Privacy Commissioner of Canada (OPC). 2023. *Announcement: OPC to investigate ChatGPT jointly with provincial privacy authorities*. [Press Release]. [Accessed July 12 2023]. Available from: https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230525-2/

Office of the Privacy Commissioner of Canada (OPC). 2021. *News Release: RCMP's use of Clearview AI's facial recognition technology violated Privacy Act, investigation concludes*. [Press Release]. [Accessed 26 June 2023]. Available from: https://www.priv.gc.ca/en/opc-news/news-and-announcements/2021/nr-c_210610/.

Risse, M. 2019. Human Rights and Artificial Intelligence: An Urgently Needed Agenda. *Human Rights Quarterly*, *41*(1), pp. 1–16. Doi: http://dx.doi.org/10.1353/hrq.2019.0000.

Salvagno, M., Taccone, F.S. & Gerli, A.G. Can artificial intelligence help for scientific writing? *Critical Care* 27(75), pp. 2-5. Doi: https://doi.org/10.1186/s13054-023-04380-2.

Scassa, T. 2022. *Comments on the Third Review of Canada's Directive on Automated Decision-Making*. [Online]. [Accessed 23 June 2023]. Available from: https://www.teresascassa.ca/index.php?option=com_k2&view=item&id=354:comments-on-the-third-review-of-canadas-directive-on-automated-decision-making&Itemid=80.

Smuha, N. A. 2021. From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), pp. 57–84. Doi: https://doi.org/10.1080/17579961.2021.1898300.

Statistics Canada. 2021. *Framework for Responsible Machine Learning Processes at Statistics Canada.* [Online]. [Accessed 25 June 2023]. Government of Canada. Available from: https://www150.statcan.gc.ca/n1/pub/89-20-0006/892000062021001-eng.htm.

Tessono, C. Yuan S., Momin M. Malik, S., Solomun D., & Sam A. 2022. *AI Oversight, Accountability and Protecting Human Rights: Comments on Canada's Proposed Artificial Intelligence and Data Act.* Cybersecure Policy Exchange. Available from: https://www.cybersecurepolicy.ca/aida.

Treasury Board of Canada Secretariat. 2023a. *Directive on Automated Decision-making.* Government of Canada. Available from: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

Treasury Board of Canada Secretariat. 2023b. List of Artificial Intelligence (AI) Suppliers. [Online]. Available from: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/list-interested-artificial-intelligence-ai-suppliers.html.

Wang, LL. 2023.Using machine learning to verify scientific claims. *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*. Organization for Economic Cooperation and Development (OECD). OECD Publishing. Paris [Online]. Available from: https://doi.org/10.1787/a8d820bd-en.

Zuiderwijk, A., Chen, Y.-C., & Salem, F. 2021. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3) 101577. Doi: https://doi.org/10.1016/j.giq.2021.101577.