

Treasury Board Directive on Automated Decision-Making

3rd Review¹

Omar Bitar, Dawn Hall, and Benoit Deshaies

Table of Contents

Executive Summary

- I. Background
- II. Purpose
- III. Current State
 - a. Algorithmic Accountability
 - b. Algorithmic Transparency
 - c. Quality Assurance
 - d. Recourse
- IV. Challenges & Risks
 - a. Overview
 - b. Scope
 - c. Periodic Review
 - d. Clients Impacted by Automated Decision Systems
 - e. Data and Model Governance
 - f. Explanation
 - g. Reasons for Automation
 - h. Peer Review
 - i. Contingency Planning
 - j. Timing of AIA Release
- V. Policy Recommendations
 - a. Effective Date
 - b. Scope
 - c. Objectives and Expected Results
 - d. Requirements
 - e. Algorithmic Impact Assessment Tool
- VI. Annex A: Proposed Amendments
- VII. References

¹ Disclaimer: This report was developed to support federal and external stakeholder consultations on the 3rd review of the Treasury Board Directive on Automated Decision-Making. The contents of the report do not represent the official view of the Government of Canada. At this stage of the review, the issues, recommendations, and provisional amendments discussed in the report only reflect the position of the Data & Artificial Intelligence Policy team members at the Treasury Board of Canada Secretariat.

Executive Summary

The 3rd review of the Treasury Board Directive on Automated Decision-Making takes stock of the current state of the policy instrument and identifies several risks and challenges to the federal government's commitment to responsible artificial intelligence (AI). It discusses critical gaps that limit the Directive's relevance and effectiveness in supporting transparency, accountability, and fairness in automated decision-making.

A key problem this review seeks to address is the Directive's scope, which excludes internal services and therefore automated decisions impacting federal employees. It also points to the need for a more holistic approach to governing data associated with AI development and use. As well, while the Directive seeks to mitigate data bias, the review notes that it does not account for bias arising from the model underlying a system.

The review determines that the Directive's approach to explanation is lacking given the lack of clarity as to what constitutes a "meaningful explanation". It also highlights the need to strengthen peer review transparency and effectiveness, as the current peer review requirement does not mandate the release of information about completed reviews or specify when they should be conducted. Similarly, the timing of the mandated release of the Algorithmic Impact Assessment (AIA) is found to be unclear in the Directive and mostly reliant on ad-hoc guidance.

In addition to identifying gaps in the instrument, the review raises issues with the terminology used in existing provisions. This is focused on the need for encompassing language to describe the subjects of automated decision-making and the importance of fostering policy coherence by aligning terms across Treasury Board policy instruments.

The rapid pace of digital transformation in the federal government has made questions around the purpose and scope of automation increasingly relevant. The review finds that the Directive does not properly consider justifications for automation or evaluate whether a system will be used within boundaries determined based on user needs and program objectives.

The review also considers broader issues related to the governance of the Directive itself. It provides an analysis showing how the current periodic review mechanism poses policy and operational challenges to the federal government.

The review concludes with policy recommendations addressing these issues. The recommendations propose to:

- **Expand the scope to cover internal services and clarify that it includes systems supporting administrative decisions;**
- **Change the periodic review interval to two years and designate a role for the Chief Information Officer (CIO) of Canada to request additional reviews, as needed;**
- **Replace references to "Canadians" with the term "clients";**
- **Introduce measures to help trace, protect, and appropriately retain and dispose of data used and generated by automated decision systems;**
- **Expand the requirement to test data for bias to cover models underlying automated decision systems;**

- **Introduce explanation criteria addressing the role of the system in the decision-making process; training and client data used; the mode of processing employed; and the outputs generated.**
- **Add a series of questions to the AIA asking institutions to provide reasons for automation;**
- **Expand the peer review requirement to mandate publication of a plain language summary and ensure that reviews are conducted prior to system deployment;**
- **Harmonize the terminology of the contingency planning requirement with that of the Treasury Board Policy on Government Security and supporting policy instruments; and**
- **Expand the AIA completion requirement to mandate publication prior to system production.**

Annex A provides provisional amendments supporting each of the policy recommendations.

Background

The Treasury Board (TB) Directive on Automated Decision-Making ('the Directive') seeks to regulate certain uses of automation technologies, including artificial intelligence (AI), in the federal public sector.² The administrative policy, which took effect in April of 2019, aims to ensure transparency, accountability, and procedural fairness in the use of automated decision systems (ADSs) to support or make administrative decisions in the federal government. Through the Algorithmic Impact Assessment (AIA), the policy adopts a risk-based approach to evaluating automation use-cases and determining the appropriate mitigations required to minimize the risk to clients and federal institutions.

The Directive targets the use of ADSs in the context of administrative decisions on external services – services the federal government delivers to external clients. Employment insurance, the Canada Emergency Response Benefit, and passport issuance or renewal are all examples of external services. The Directive has highlighted the importance of responsible AI and prompted many federal institutions and public sector organizations in other jurisdictions to draw on it as a best practice in order to identify and address the risks of automating administrative or other decision-making processes.

The Treasury Board of Canada Secretariat (TBS) is obliged to conduct a review of the Directive every 6 months to ensure that it remains relevant and responsive to the evolving AI landscape in the federal government. The 1st review led to a series of amendments to the Directive that mainly sought to clarify and reinforce existing requirements, update policy references, and strengthen transparency and quality assurance measures. The 2nd review informed the development of guidelines supporting the interpretation of the Directive. This report constitutes the 3rd review of the Directive.

² The Directive can apply not just to AI systems, but to any automated decision system, a category that encompasses a wide range of digital systems with varying degrees of complexity – including those that possess AI capabilities. Under Appendix A of the Directive, automated decision systems are broadly defined to cover “techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.” The title of the Directive reflects the government’s intent to regulate the automation of a specific type of decision, regardless of the choice of technology.

In the Foundation Framework for Treasury Board Policies, a Directive is a type of mandatory policy instrument usually addressed to managers and functional specialists. Directives provide “formal instruction that obliges departments to take (or avoid) specific action.” They “explain how deputy heads' officials must meet” a policy objective.

Since the Directive was issued, various groups including ministers' offices, academics, and journalists have expressed concerns with the scope of the Directive. Revelations of AI use in hiring processes in the government have further amplified calls for reconsideration of the scope, which falls short of protecting federal public servants. These developments have informed the focus of the current review, which highlights what's at stake for federal employees subject to automated decision-making without adequate governance and oversight. This is one of several issues considered in this review, which also identifies other gaps in the Directive and highlights issues with the feasibility, terminology, and coherence of existing policy requirements.

Purpose

The purpose of this report is to satisfy the Directive's periodic review requirement. Serving as the 3rd review of the instrument since its release, the report discusses the current state of the Directive and identifies several issues that challenge the federal government's commitment to responsible AI and limit the Directive's effectiveness in achieving its objectives and expected outcomes. The report proposes potential approaches to addressing the issues and concludes with policy recommendations intended to guide the evolution of the Directive. Periodic reviews are not intended to be exhaustive. Instead, they seek to identify and address issues which have been prioritized due to their urgency or impact.

Current State

Overview

The scope of the Directive is among its distinguishing features. It is not designed to regulate any use of AI in the federal government; the instrument is rather concerned with the use of ADSs in administrative decision-making. ADSs are defined broadly to include "any technology that either assists or replaces the judgment of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets." The Directive applies to ADSs used to render or inform administrative decisions on external services, which the TB Policy on Service and Digital defines as services "where the intended client is external to the Government of Canada." The Directive defines an administrative decision as any "decision that is made by an authorized official of an institution as identified in section 9 of this Directive pursuant to powers conferred by an Act of Parliament or an order made pursuant to a prerogative of the Crown that affects legal rights, privileges or interests."

The Directive is issued under the authority of the *Financial Administration Act* (FAA) and applies to departments as defined in section 2 of the Act. It is rooted in administrative legal principles such as transparency, accountability, legality, and procedural fairness. It extends well-established principles of human decision-making into the evolving realm of automated decision-making. It is also a supporting instrument under the TB Policy on Service and Digital, which aims to integrate the management of data, information, information technology (IT), cyber security, and service in the Government of Canada (GC). These foundations serve as anchors for the Directive, informing the ways in which it is applied to use-cases involving decisions that affect the legal rights, privileges, or interests of an individual or business.

The requirements of the Directive can be grouped into four thematic categories:

- (a) Algorithmic accountability;**
- (b) Algorithmic transparency;**
- (c) Quality assurance; and**
- (d) Recourse.**

Algorithmic accountability

The Directive establishes unique accountabilities for the use of ADSs in the federal institutions bound by it. It designates Assistant Deputy Ministers (ADMs) responsible for the program using an ADS (or any other official named by the institution's Deputy Head) as responsible for complying with the requirements of the Directive. This strengthens accountability for AI, facilitates compliance, and fosters responsible AI practices across the GC.

The Directive's parent policy, the TB Policy on Service and Digital, holds Deputy Heads accountable for "ensuring the responsible and ethical use of automated decision systems, in accordance with TBS direction and guidance". This requires Deputy Heads to ensure that decisions produced by ADSs meet the criteria of the Directive (e.g., efficient, transparent, explainable, unbiased); disclose information about the use of AI systems in scope, including through the AIA tool; and conduct ongoing risk assessment and management in accordance with the risk mitigation measures appropriate for the impact-level determined by the AIA.

These accountabilities cover both ADSs and the data collected for and used by them; the Directive includes requirements pertaining to both. However, as an increasing number of federal institutions establish departmental data governance structures and clarify roles and responsibilities for data, additional clarity may be needed to delineate data and AI accountabilities.

The Directive also identifies responsibilities for responsible AI at the enterprise level. It holds the Chief Information Officer (CIO) of Canada responsible for providing guidance on the use of ADSs, granting exemptions, developing and maintaining the AIA tool, and communicating and engaging with federal and other partners on responsible AI.

The consequences of not complying with the requirements of the Directive and other mandatory policy instruments are established in the Framework for the Management of Compliance. Appendix C and D of the framework lay out possible consequences for institutions subject to such instruments and individuals responsible for complying with them. The consequences are classified by degree of severity, which range from the minimal to the most severe. Different types of corrective actions are reserved for each degree of severity. For example, minimal consequences of noncompliance could include persuasive measures such as training, education, or collaboration. In more severe cases, noncompliance could lead to financial penalties for individuals and constraints on authorities for institutions. While TBS has not taken any disciplinary actions in the case of the Directive, the framework can help assure that institutions and designated officials are accountable for automation projects impacting external clients.

Algorithmic transparency

Algorithmic transparency is one of the Directive's organizing principles. Institutions are required to complete and publish an AIA to the Open Government Portal. Developed in the open, the AIA covers a

broad range of topics related to the system's design, algorithm, decision type, impact, and data. The questionnaire evaluates the impact of an ADS on the rights, health and well-being, and economic interests of individuals. It also addresses systemic risks by considering impacts on the sustainability of ecosystems. There are four levels of impact (I-IV, little to no impact to very high impact), each with corresponding risk mitigation measures which specify or add to the Directive's requirements.

Since the Directive took effect on April 1st, 2019, five AIAs have been released on the Open Government Portal. The low count can be attributed to the Directive being only applicable to ADSs developed or procured after April 1st, 2020. During this period, many departments and agencies had to reallocate resources to support the government's response to the COVID-19 pandemic. It is also possible that many federal institutions had not been aware of the Directive or informed about the types of cases it is designed to regulate. The rate of AIA completion and release will likely increase as more automation projects fall within the purview of the Directive and federal institutions become increasingly aware of the purpose and scope of the instrument.

Once the ADS is in production following the completion of an AIA, effectiveness and efficiency data must be reported. In addition, the full or partial automation of a process leading to an administrative decision (impact level II or higher) must be clearly noted on relevant websites or other service delivery channels. This is complemented with an obligation to provide meaningful explanations of decisions to clients (the Directive does not define what constitutes a meaningful explanation). Explanation requirements vary by impact level. Impact levels II-IV require explanations for decisions resulting "in the denial of a benefit, a service, or other regulatory action" (Appendix C).

The problem of so-called 'black box' algorithms highlights the importance of designing and using ADSs that can produce explainable outputs. Explainability, however, doesn't guarantee interpretability, without which explanations would not be meaningful to clients or users. The distinction between explainability and interpretability is subtle but significant, especially given the broad range of technologies with varying levels of complexity covered under the definition of an ADS, which states that the systems "draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets." The TB Guideline on Service and Digital ('the Guideline') distinguishes between explainability and interpretability, noting that the former involves techniques that help approximate how a model produces an output while the latter concerns the extent to which this process can be communicated and understood (subsection 4.5.3). The Guideline recommends that federal institutions factor in considerations of explainability and interpretability when selecting machine learning models. The higher the complexity of a model, the more difficult it may be for users to provide meaningful explanations that are interpretable to themselves and to clients. The Guideline cites neural networks and deep learning as examples of techniques that may pose challenges in this regard. Peer reviewers, legal services, and system developers have a significant role to play in determining whether the choice and design of an ADS can satisfy the Directive's explanation requirements.

The GC's ability to access and evaluate proprietary or government-owned software is helpful for determining whether the federal institution seeking to use it can comply with the Directive effectively. We may find, for example, that some highly elaborate technologies operate opaquely and could

complicate the release of data or documentation required under the Directive (e.g., in support of the notice or explanation requirement). Requiring federal institutions to guarantee access to proprietary ADS software and release any non-sensitive custom source code is another mechanism through which the Directive helps foster algorithmic transparency. The assessment of a system's potential to meet the requirements of the Directive can also take place during the procurement phase. The AI Source List provides federal institutions with over 100 prequalified vendors that would be capable of meeting the Directive's requirements should they be contracted to provide the government with AI solutions. The Guideline highlights the value of engaging with vendors while completing an AIA in order to better understand the risks of an ADS and plan for the implementation of mitigation measures.³

Quality assurance

Data is the fuel for AI. The outputs and outcomes of algorithmic technologies significantly depend on the quality of the data used to train and optimize them. The Directive dedicates a set of requirements to 'AI data'. This includes assessing the quality of data prior to the production of the system in order to minimize bias and ensure fitness-for-purpose. It's noteworthy that the requirements specifically address input data — data collected for and used by the system.

During the planning and development stages, legal consultations are required to evaluate the legality of an automation project. This could involve consultations with an institution's Access to Information and Privacy (ATIP) office to evaluate compliance with applicable privacy policy and legislation. Seeking input from centers of expertise for other government-wide considerations such as accessibility or official languages can also support compliance with applicable laws and policies. Institutions are also required to conduct security risk assessments during system development and to establish the appropriate safeguards in accordance with the TB Policy on Government Security. The language of ethics in the TB Policy on Service and Digital (subsection 4.4.2.4) supplements such checkpoints, helping ensure that AI use is both legal *and* ethical. Once in production, systems must be continuously monitored to ensure that they are compliant with applicable policy and legislation. This supports the reporting requirement noted above. Determining whether an ADS meets these standards requires policy, legal, ethical, and technical expertise.

The Directive's peer review requirement for systems at impact levels II-IV provides a mechanism for external validation by the appropriate experts from government, academia, industry, non-governmental organizations, and/or third parties with the relevant areas of specialty. This not only helps validate the risks that these systems might pose to clients, but also guards against conflicts of interest by drawing on external sources of expertise. To avoid disruptions to processes relying on ADSs (impact levels III-IV), the Directive requires users to have a contingency plan in place should an ADS become unavailable. Disruptions could arise from technical difficulties, emergencies (e.g., as a result of pandemics, natural disasters, or cyber-attacks), or decisions to suspend or modify an automation project.

For high and very high impact systems (levels III-IV), the Directive requires human intervention; that is, humans to oversee the process and make the final decision. The importance of including 'humans in the loop' in automated decision processes points to the value of building AI and data literacy across the GC. The Directive accounts for this by requiring training to build employees' capacity to review, explain, and

³ The role of economic and geopolitical factors (e.g., obligations under free trade agreements, risks to technological sovereignty) in shaping decisions on technology procurement will have to be explored as the GC matures its approach to responsible AI.

oversee ADS operations. While the involvement of human decision-makers can help mitigate ADS risks in high-impact scenarios, however, human bias towards automated outputs may interfere with this process. Known as automation bias, this is a tendency to perceive outputs of algorithmic technologies as objective or neutral.

Recourse

Algorithmic transparency and accountability would be incomplete if automated decisions were final or irreversible. The Directive establishes a recourse mechanism through which clients may challenge an administrative decision made or informed by an ADS. This is key to building public trust in an increasingly digital government and maintaining the democratic legitimacy of federal institutions. The recourse mechanism can also help reduce the need to resort to judicial review of automated administrative decisions subject to the Directive.

While it enables clients to challenge administrative decisions *ex-post*, the Directive does not include measures allowing clients to opt out of automated decision-making. This type of requirement may become increasingly important as more government services are digitalized end-to-end and automated. If the scope of the Directive were to expand to cover a wider range of administrative decisions, opt-out requirements may arguably become desirable for certain use-cases. Article 22 of the EU's General Data Protection Regulation (GDPR) provides an example of a consent-based opt-out model: it empowers citizens with a right to opt out of being subject to fully automated decisions involving personal data processing with significant legal impacts. Legislative measures may be needed to establish or recognize an equivalent right at the federal level. Research and consultation with stakeholders will be needed to determine the government's direction in this regard.

Challenges & Risks

Overview

New types of AI use-cases in the GC and ongoing institutional engagement with the text of the Directive have enabled us to identify various gaps and issues in this review. TBS's internal governance of the Directive has also produced lessons on the challenge of balancing stability with adaptability in such a dynamic and constantly evolving space. This section distills key issues accumulated while supporting the implementation of the Directive over the last two years. Informed by the current state presented in the previous section, it provides an analysis of the risks and challenges each issue poses to the Directive's effectiveness as a policy and to the government's broader commitments to responsible AI in the GC and internationally. Some issues (e.g., on scope limitations) respond to external developments while others (e.g., inconsistent or confusing terminology) are internal to the text of the instrument and to Treasury Board policy more generally. The issues can be broken down and described as follows:

- **Scope:** Current scope excludes internal services and is expressed in potentially confusing terms.
- **Periodic review:** 6-month review interval presents policy and operational challenges to TBS.
- **Clients impacted by automated decision systems:** References to Canadians fail to recognize other service recipients impacted by ADSs.

- **Data and model governance:** There are no dedicated data governance measures to ensure that data used and generated by ADSs are traceable, protected, and retained and disposed of appropriately. There are also no provisions addressing model bias.
- **Explanation:** Current requirement does not define what constitutes a “meaningful explanation”.
- **Reasons for automation:** There are no mechanisms to provide clients and public stakeholders with a justification for a program’s decision to adopt AI and a description of the boundaries for its use in relation to user needs and program goals.
- **Peer review:** The lack of a requirement to publish information about peer reviews is a missed opportunity for the government. It is also unclear when reviews should be conducted.
- **Contingency planning:** The terminology used to express contingency planning requirements does not align with relevant language used in the TB Policy on Government Security and supporting policy instruments.
- **Timing of AIA release:** AIA requirements do not specify a timing for release, creating uncertainty as to when an AIA should be published and potentially weakening the Directive’s transparency measures by allowing institutions to delay AIA release well into a system’s lifecycle.

This is not intended to be an exhaustive list of problems with the Directive. It is only illustrative of the key issues that TBS has grappled with in the course of overseeing the Directive and the AI policy space in the GC. TBS recognizes that more remains to be done even after all these issues are duly addressed.

Scope

Recent domestic and global trends in AI use to hire, monitor, track, and evaluate employees suggest that the scope of the Directive merits reconsideration. In the public and private sectors, algorithmic technologies have been deployed as tools of surveillance and gamification, raising ethical and legal questions about the appropriate balance between optimizing organizational efficiency and preserving the dignity and autonomy of workers. The use of automated systems to enhance the provision and management of human resources also presents challenges to employers’ ability to fulfill their obligations towards prospective candidates and employees. The risk of introducing unintended biases into hiring or promotion processes, for example, could undermine an employer’s commitment to uphold employment equity or adhere to established labour standards. Such risks have prompted some governments to regulate the use of AI in employment contexts.

Last year, New York City passed a landmark law regulating the use of “automated employment decision tools” used to support or replace discretionary decision-making in hiring or promotion processes. The law establishes safeguards against bias and measures supporting algorithmic transparency and accountability. It requires employers to conduct bias audits on the tools they plan to use; notify candidates or employees of automation in decision-making processes and allow them to request alternative selection processes; and disclose assessment criteria. The law also empowers employees or candidates to request information about data collected for an automated decision tool. Noncompliance with the law could make employers liable to civil penalties. The New York City law is one of several initiatives in the United States seeking to address the use of AI in employment contexts. Illinois and Maryland have passed similar laws regulating the use of AI in virtual job interviews.

At the federal level, the US Equal Opportunity Employment Commission launched in 2021 an initiative to ensure that the use of AI in employment settings does not violate federal civil laws. The initiative aims to

guide applicants, employers, employees, and vendors to foster fairness in the use of algorithms in accordance with federal equal employment opportunity laws. In the United Kingdom, the all-party parliamentary group on the future of work has recommended an “accountability of algorithms act” that would establish safeguards for AI used to surveil workers or optimize their performance. The proposed legislation would establish a public sector and corporate duty to “undertake, disclose and act on pre-emptive Algorithmic Impact Assessments” for automation impacting individuals in the workplace.

The Directive was not designed to address AI risks impacting individuals within the federal public service. It does not cover internal services, which the Guideline defines as “groups of related activities and resources that the Government of Canada considers to be services in support of programs or required to meet corporate obligations of an organization.” In the GC, human resource management is considered a type of internal program associated with services such as employee staffing, recruitment, and compensation. Given their legal significance, determinations in these areas can be considered administrative decisions.

Recent media reports in Canada have shed light on the use of algorithmic technologies in the GC to enhance hiring processes. As part of a recruitment campaign to improve diversity and inclusion within executive ranks in the federal public service, automated systems were used to shortlist candidates based on behavioral assessments and match them with vacancies based on their compatibility with preset criteria. The use of ADSs in hiring contexts could also support application screening, analysis of candidates during online interviews, and ranking of candidates based on a score or other evaluation scheme. Due to unintended biases, high error rates, or opaque outputs, the use of AI to help filter or select job candidates could risk undermining the GC’s commitment to fair and impartial hiring processes and to a diverse, equitable, and inclusive workforce. The exclusion of internal services from the scope of the Directive thus risks damaging public trust in the government’s capacity to strike an appropriate balance between innovation and responsibility.

The Institute for the Future of Work (IFOW) has examined the use of AI in hiring, recruitment, promotion, and terms of work. Researchers at the institute state that the outputs of machine learning techniques used in these areas are based on correlations between representations of a candidate or employee and a corresponding ideal. The challenge is that this ideal is likely to be constructed from biased historical data. Amazon’s experiment with AI use in hiring demonstrates this: the company realized that their system was exhibiting a gender bias because it was trained using data from résumés submitted over the previous decade. Most of the résumés were from men, thus biasing the system to prefer applications with male attributes and devalue qualifications associated with résumés belonging to women.

Even representations of an applicant can be inaccurate or unreliable, for example in the case of facial or affect recognition tools. The error rates observed in such machine learning technologies can create ethical and legal challenges for users looking to use them to evaluate candidates during online interviews or monitor the performance of employees, as they can lead to discriminatory outcomes and perpetuate inequalities. The involvement of a human decision-maker in such scenarios would not necessarily eliminate these risks, as algorithmic outputs (especially in the presence of an automation bias) can distort the judgment of human decision-makers.

The significance of decisions supporting human resources management in the GC underlines the importance of ensuring that the appropriate safeguards are in place to minimize the risks of automation

to the rights, privileges, and interests of federal employees. Expanding the scope of the Directive to cover internal services would accomplish this goal. The value of an expanded scope would not be limited to the human resources context. The GC Strategic Reference Model identifies 10 internal program types, each of which houses a number of internal services. Other services that would likely benefit from being subject to the safeguards of the Directive include grants and contributions, awards and recognition, and security screening. It's worth noting, however, that not all internal services involve administrative decision-making. Internal enterprise services, in which the provider and recipient are federal institutions, are unlikely to become subject to the Directive except where they specifically concern the rights, interests, or privileges of individual employees.⁴

In addition to being narrow, the scope of the Directive is articulated in potentially misleading terms. Section 5.2 of the Directive states that it applies to “any system, tool, or statistical models used to recommend or make an administrative decision about a client.” The distinction between recommending and making a decision is unclear and has caused confusion about the conditions under which ADS use is subject to the Directive. The challenge is that the term ‘recommend’ appears to set a high bar for ADS involvement in a decision-making process, thus leading to the conclusion that less significant degrees of involvement are outside the bounds of the Directive. The intent, however, is to cover any system with the potential to influence an administrative decision-making process. Whether it is recognizing text in client documents or generating a score to inform an eligibility decision, an ADS is within scope. This problem can be remedied by replacing the term ‘recommend’ with a term such as ‘support’, which indicates a more generic function. This would better distinguish between systems that make a decision and systems that do not, while clarifying the Directive’s potential to cover both.

Periodic review

The Directive’s biannual review mechanism is intended to ensure that the instrument remains relevant and responsive to the evolving risks and challenges of automated decision-making in the federal public sector. The value of this mechanism has been demonstrated in previous reviews of the Directive, which culminated in improvements to the text of the instrument. The first review, for example, clarified multiple requirements and strengthened transparency and quality assurance measures. The review mechanism also serves as an opportunity to engage stakeholders in the federal government, other sectors, and the wider public on the regulation of ADSs in the GC. In this way, it helps foster a multi-stakeholder and participatory policy-making process.

⁴ The TB Policy on Service and Digital defines an internal enterprise service as “a service provided by a Government of Canada department to other Government of Canada departments intended on a government-wide basis.” Under section 5 of the Policy, the roles and responsibilities laid out for Shared Services Canada (SSC) and Public Services and Procurement Canada (PSPC) provide examples of internal enterprise services. SSC is responsible for providing “certain services related to email, data centres, networks and end-user technology devices” to other federal institutions, while PSPC is responsible for providing “services for federal departments and agencies, to support them in the achievement of their mandated objectives as their central purchasing agent, linguistic authority, real property manager, treasurer, accountant, integrity adviser, and pay and pension administrator.” Some of these services would not fall within the purview of the Directive as they are delivered to enable the operations of an organization and do not specifically concern individual employees. Decisions in such contexts are therefore unlikely to be considered administrative. Nevertheless, TBS plans to engage with internal legal services and colleagues within the Department of Justice to better understand how best to evaluate and determine the legal status of internal service decisions.

Despite all its merits, the current periodic review requirement presents significant operational challenges. First, the mandated 6-month period may not be suitable if we consider the pace at which the federal AI landscape has evolved since the Directive was issued. While the COVID-19 pandemic may have accelerated the use of AI systems across government, the adoption of automated systems to make or support administrative decisions has been relatively slow. This is evidenced by the number of AIAs published on the Open Government Portal. (It's worth reiterating that the Directive only applies to systems developed or procured starting April 1st, 2020.) Generally, it takes time to identify and validate gaps in policy or issues impacting effective compliance. This is partly dependent on the evidence base, which crucially relies on departments and agencies interacting with the Directive by evaluating its applicability and adopting its measures. It is also dependent on the wider federal and international AI landscape, where legal or policy developments can expose gaps in the Directive and influence its long-term trajectory. Second, it is important to consider the impact of frequent changes to policy on predictability, an important value in administrative rule-making. As with any administrative policy, the requirements of the Directive should display a degree of stability and reliability, enabling federal institutions and the clients they serve to plan and act with a reasonable degree of confidence. On this view, the biannual review period may be an obstacle given the uncertainty it creates for the users and subjects of ADSs.

Third, amending Treasury Board Directives can be a lengthy process, not only because of the extensive consultations involved but also due to the level at which they are approved. The amendments proposed as part of the first review of the Directive took over 18 months to finalize. By the time the updated Directive was published following the approval of the Secretary of the Treasury Board, the federal AI landscape had significantly changed. This mismatch between policy and governance timelines risks creating a situation whereby the results of reviews are outdated or no longer relevant by the time they are released. Fourth, the frequency of the periodic review requirement places a significant burden on the team overseeing the Directive. The capacity and time needed to undertake a full review (especially in cases involving amendment proposals) demand the continuous allocation of resources to this effort. In practice, what this means is that at any point during the fiscal year, the team is likely to be in 'review mode', with one or even two reviews underway. This limits the team's ability to support the implementation of the Directive through awareness raising, targeted outreach, or interpretive guidelines.

Taken together, these challenges can be addressed by extending the interval between reviews. Instead of a 6-month period, a more flexible approach requiring at least one full review every two years could provide policymakers with more time to take stock of the strengths and weaknesses of the Directive, while enabling them to better divide resources between policy review, oversight, and implementation support. Given their responsibilities under the TB Policy on Service and Digital, the CIO of Canada can also play a role in requesting the completion of off-cycle reviews of the Directive to address urgent issues or priorities.⁵

⁵ Subsection 4.1.2.1 of the Policy states that the GC CIO is responsible for, *inter alia*, advising the Secretary and President of the Treasury Board on the use of "emerging technologies and the implications and opportunities of doing so for the Government of Canada".

Clients impacted by automated decision systems

The Directive uses several terms to refer to subjects of automated decision-making. The objective in section 4.1 refers to Canadians, while the scope described under section 5.2 refers to clients more broadly. This poses an issue not only due to the inconsistency in the terminology, but also because of the problems we encounter when singling out Canadians in the text of the instrument. Administrative decisions made within the federal government do not only affect Canadian citizens; they can also impact permanent residents, asylum seekers, or residents with a student visa or work permit – among other categories of citizenship status. They can also have extraterritorial effects to the extent that they impact foreign nationals abroad, as observed with decisions on visa applications or sanctions targeting foreign actors.

Given the broad range of areas in which government officials make decisions of an administrative nature, the potential beneficiaries of the instrument's safeguards and protections are not limited to Canadian citizens. Thus it would be more appropriate to refer to clients in the objective statement, while ensuring consistency throughout the text (and in the phrasing of questions in the AIA). This would also be more inclusive, encompassing a wider range of individuals and businesses subject to federal jurisdiction.⁶

Data and model governance

The Directive currently dedicates requirements to what can be called input data; that is, data collected for and used by an ADS. Before deploying an ADS, institutions are required to establish processes for testing data used by the system for unintended bias. The quality of this data, which could include training or client data, must be validated to ensure it meets quality criteria such as accuracy, relevance, and timeliness. The AIA tool supports these requirements with questions about who controls this data and whether it contains personal information. These measures enable compliance with the TB Policy and Directive on Service and Digital, and with the *Privacy Act* and supporting TB privacy policy.

What is missing from the Directive is a more holistic approach to the governance of data — not just data inputs, but also outputs generated by a system, which could include recommendations, scores, rankings, or inferences. In a paper on data governance, the Global Partnership on Artificial Intelligence (GPAI) Working Group on the Responsible Development, Use and Governance of AI identifies multiple facets of data governance for “AI training and testing data”, “algorithmic input and output data”, and “wider data ecosystems”. It lays out considerations for each category of data. The Directive already accounts for some issues such as quality and bias, which are identified for data collected for and used by systems. Other issues identified for both data inputs and outputs, such as data protection, safeguards to minimize the risk of harm, and data source transparency, are not specifically addressed in the instrument.

Supporting data traceability can strengthen transparency around the provenance of data inputs and outputs. A program's ability to provide adequate notice and explanation to clients partly depends on this data being traceable to their source. Readiness for litigation or audit may also require this. The Directive

⁶ Under the TB Policy on Service and Digital (Appendix A), a client is defined to include “Individuals, businesses or their representatives served by or using either internal or external services provided by the Government of Canada. When describing interactions with information technologies, clients can be referred to as users.”

recognizes a comparable need in subsection 6.2.8, which requires the documentation of decisions.⁷ The AIA also includes a question on whether an ADS can generate an audit trail of all previous recommendations or decisions.

Data protection is also critical, as both inputs and outputs could include personal information or lead to personally identifiable information in combination with other data. Establishing safeguards to protect the privacy and security of this data is key to ensuring that it is not shared or reused inappropriately. Given the scope of the Directive, system outputs are likely to include potentially sensitive information which, if leaked, disclosed, or otherwise mismanaged, could render clients vulnerable to discrimination or abuse. The protection of this data would support security safeguards required under subsection 6.3.7, and support adherence to privacy and security policy and legislation at large.

Programs running an automation project should also establish retention and disposition schedules for this data to ensure it is not held longer than is necessary or disposed of without regard to relevant disposition authorities. The TB Directive on Service and Digital requires institutions to establish retention periods and documented disposition processes for all information and data in accordance with the *Privacy Act* and Library and Archives Canada's disposition authorizations (subsections 4.3.1.7, 4.3.1.8). Integrating retention and disposition considerations into the design and implementation of an ADS supports compliance with these instruments, helps mitigate the risk of harm, and contributes to building public trust in the use of AI in the federal public sector.

Achieving the desired outcomes of the Directive may require governance not only of input and output data, but also of models underlying ADSs. Biased outputs or decisions can arise either from the data used to train an ADS or the model processing input data. Model-related biases are not addressed in the Directive's requirements, though they are arguably implicit in subsection 6.3.2, which requires monitoring of system outcomes. Monitoring, however, occurs following deployment. What is needed is pre-production testing for model bias. Subsection 4.5.3 of the Guideline speaks to the problem by identifying factors institutions should consider during model selection. The Guideline recommends that institutions opt for "the simplest model that will provide the performance, accuracy, interpretability and lack of bias required." Closing this gap in the Directive will help ensure that key sources of bias are tested for and mitigated prior to system production.

Explanation

The current explanation requirement mandates federal institutions using ADSs to provide "meaningful explanations" of how and why an automated decision was reached. While the Guideline (subsection 4.5.3) clarifies the meaning of "explainability" to inform decisions related to model selection, the requirement is still amenable to many interpretations. Appendix C of the Directive requires general explanations for low impact "common decision results" and tailored explanations for decisions at impact levels II-IV that "resulted in the denial of a benefit, service, or other regulatory action". But it does not prescribe criteria

⁷ It's unclear if the requirement, which concerns "decisions of Automated Decision Systems", covers data produced by the system (data outputs) and used to inform a decision made by a human. Ensuring that data outputs are traceable whether or not they constitute an administrative decision can address any uncertainties surrounding such cases of partial automation.

identifying information that must be submitted as part of an explanation of a decision at a certain impact level. This creates several problems for federal institutions, TBS policy leads, and clients.

First, varying interpretations could result in inconsistent practices, which may lead to incomplete explanations and disparities in the treatment of clients. Institutions that generously interpret the phrase “meaningful explanation”, for example, may be inclined to provide clients with more comprehensive documentation describing how a decision was reached than ones adopting a narrow reading. In general, vagueness in administrative policy creates legal risks for governments, as it comes into tension with core administrative legal principles such as legality and procedural fairness.

Second, the lack of clear direction on what constitutes a “meaningful explanation” can lead to an ad-hoc approach to explainability, with federal institutions seeking guidance from TBS policy leads on a case-by-case basis. This is not only burdensome for policymakers overseeing the Directive, but also perpetuates an environment of uncertainty and complicates compliance.

Explainability is a common feature in AI policy and regulatory initiatives worldwide. The attempt to afford subjects of automated decisions with legible information explaining how algorithmic outputs impacting them were attained has taken various forms including rights, ethical principles, and policy provisions. Like the Directive’s explanation requirement, many measures in other jurisdictions do not specify what information would constitute an adequate explanation. France’s 2016 *Loi pour une République numérique (Digital Republic Act)* stands out in this regard: in addition to enshrining in law a right to explanation for automated administrative decisions, it lays out an explainability framework detailing the information organizations would need to provide when asked to communicate the rules and characteristics underlying an automated system:

- “1. The degree and the mode of contribution of the algorithmic processing to the decision making;
2. The data processed and its source;
3. The treatment parameters and, where appropriate, their weighting, applied to the situation of the person concerned; and
4. The operations carried out by the treatment.”⁸

The Directive already requires some of this information under certain conditions, though the goal is notification rather than explanation. For ADSs at impact levels III-IV, institutions are required to notify clients of how the components of the ADS work; how it supports the administrative decision; what training data was used; and the results of any audits or reviews of the system. While this requirement can in some cases bolster the “meaningful explanation” provided to a client following a decision, it doesn’t address the gap identified in the explanation requirement.

Expanding the explanation requirement with criteria similar to those identified in the French law can clarify the meaning of the requirement and help ensure consistent interpretation. It would also better position the Directive to account for the digital character of automated decisions, which demands unique measures targeting the algorithm, data, and other components to ensure that government institutions

⁸ These elements were added by decree in 2017 to elaborate the right to explanation initially established in the law.

looking to augment or replace human decision-makers can continue to meet the standards of administrative law.

Explanation criteria would respond to the need to safeguard a client’s right to a fair and impartial decision-maker, and to reasons for decisions impacting them. This requires programs running automation projects to be able to supply information on the design and implementation of an ADS as well as the role it played in a decision-making process. Whether a decision is being partially or fully automated, the process by which a system arrives at or contributes to a decision must be interpretable for users and clients alike. This requires clear descriptions not only of the mode of processing employed by an ADS, but also of the parameters (e.g., business rules) used to evaluate client data. The output and its relationship to the final decision must also be described to clients. These considerations can be summed up as criteria:

- The role of the system in the decision-making process;
- The training and client data, their source and method of collection, if applicable;
- The criteria used to evaluate client data and the operations applied to process it; and
- The output produced by the system, and any relevant information needed to interpret it in the context of the administrative decision.

Embedding these criteria in Appendix C of the Directive (under the explanation requirement section) would help create consistency in the explanations offered to subjects of automated decision-making, reducing legal risks to the government and to clients. Federal institutions could also leverage the criteria to inform decisions on ADS procurement or development, recognizing that not all AI models are equally explainable.

Reasons for automation

One of the guiding principles for responsible AI in the GC emphasizes transparency and the value of “starting with a clear user need and public benefit”. The ongoing adoption of AI technologies in public sector institutions has not always been accompanied by a clear and demonstrable need. Rather, the popularity of automation as an innovative solution has made it an attractive option across business lines. In a briefing on the state of the federal AI landscape during the 2019-20 fiscal year, Shared Services Canada found that around 80% of federal institutions were exploring AI, and documented over 300 AI projects at varying stages of development in the government. As automated systems are mainstreamed into the operations and decision-making processes of government departments and agencies, justifying the decision to procure or develop such systems becomes crucial for informing users and stakeholders about their necessity for fulfilling a mandate or achieving a goal.

Even where AI technologies are deemed necessary — to help process growing volumes of applications or monitor epidemiological trends, for example — it is also important to set clear boundaries for their use in proportion to the needs they are addressing. The Office of the Privacy Commissioner of Canada’s “privacy principles for contact tracing and similar apps” provide an example of how principles of necessity and proportionality can be applied in contexts involving the design and deployment of technologies with potentially significant impacts on individual privacy and autonomy. Such projects, the guidance states, must be “necessary for a specific purpose, tailored to that purpose and likely to be effective.” It describes necessity in terms of the “purpose or purposes underlying a measure”, which “must be evidence-based

and defined with some specificity”. Proportionality requires that a measure “be carefully tailored in a way that is rationally connected to the specific purpose(s) to be achieved”. While these principles were promulgated in an effort to guide the development and release of the COVID Alert mobile application and other technological measures proposed to help combat the spread of COVID-19 across Canada, they might also be useful in informing how institutions justify their decision to pursue automation.⁹

The Directive does not ask users to explain their decision to leverage AI technologies in administrative decision-making. While the AIA includes questions asking users to describe their automation project and identify relevant business drivers (e.g., lower transaction costs, improve the quality of a decision), users are not expected to provide reasons justifying the necessity of automation for meeting specific user needs. Similarly, the Directive and AIA do not account for whether the use of an ADS will be confined to meeting user needs and program objectives. Nevertheless, articulating the purpose and scope of an automation project can play an important role in the assessment of automated decisions, for instance in the context of judicial review. Here, this information could provide auxiliary considerations supporting analyses of whether an institution running an automation project adequately respected applicable procedural rights. There is value in making such considerations more salient in policy, not necessarily in the form of a distinct requirement (it would be difficult to define and consistently apply general criteria in this area) but rather as questions for project leads to reflect on and address. The AIA can provide a space for institutions to provide a rationale for an automation project, describing not only why it’s necessary but also whether it is compatible with user needs and appropriate for program objectives. This information would be openly available to federal and public stakeholders, creating new opportunities for cross-sectoral dialogue on the merits of automation and the limits of ADS use in administrative decision-making.

Peer review

The peer review requirement provides an additional layer of assurance to system developers, users, and subjects of ADSs. The evaluation of a system by a subject matter expert within or outside the GC has the potential to identify new risks in a system, corroborate the claims of the project lead prior to or following the implementation of risk mitigation measures, and drive cross-sectoral collaboration.

Peer review also has the potential to strengthen algorithmic transparency and accountability. It can inform the ADM responsible for the automation program of any residual ADS risks which haven’t been effectively addressed through applicable mitigation measures. This can support the ADM’s assessment of the overall project and help determine whether they should approve it. Peer review can also provide members of the public with an assessment of the performance of a system and the guardrails in place to ensure compliance with the Directive and minimize risks to clients and federal institutions.

The current peer review requirement does not mandate the release of peer reviews or related information. This creates a missed opportunity for bolstering public trust in the use of ADSs through an externally sourced expert assessment. Given that the AIA is completed and released by project leads, public stakeholders and clients subject to automated decision-making could benefit from external

⁹ Some federal institutions have also considered how principles of necessity and proportionality could inform assessments of data collection and use (see Statistics Canada’s Necessity and Proportionality Framework) and privacy impacts (see the Office of the Privacy Commissioner of Canada’s Guide to the Privacy Impact Assessment Process).

validation of their characterization of a system. (Another source of publicly available information about ADSs can be found in vendor “Invitation to Qualify” submissions, which are published on Public Services and Procurement Canada’s website.)

Peer reviewers can be bound by non-disclosure agreements when evaluating proprietary source code or trade secrets. Similarly, information related to AI systems built in-house could be classified. Making peer reviews publicly available, therefore, may not always be feasible. Such limitations can be addressed by requiring at least a *summary* of a peer review. The summary could describe the scope, purpose, and findings of a review, while laying out any recommendations to the relevant federal institution. Mandatory publication can be built into the peer review requirement (subsection 6.3.4). This would apply to automation projects at impact levels II-IV. Published peer review summaries would complement other documentation the Directive requires project leads to disclose for projects at any of these levels. This includes information about how the components work; how the system supports the administrative decision; results of any audits or other reviews; and descriptions of (or links to) training data.

Another gap in the peer review requirement concerns the timing of review. The current text does not specify when project leads should engage their selected reviewers to undertake a review, creating uncertainty for both institutions and reviewers as to whether to complete a review prior to or during system deployment. Unlike audits, reviews are most effective when made available alongside an AIA, prior to the production of a system, so that they can serve their function as an additional layer of assurance. Specifying that peer reviews should take place prior to the production of an ADS would address this problem.

Contingency planning

Subsection 6.3.6 of the Directive directs institutions to establish backup systems and/or processes to ensure service continuity, particularly for automation projects at impact levels III-IV. The requirement supports compliance with federal security policies such as the TB Policy on Government Security (PGS) and TB Directive on Security Management (DSM). The PGS (subsection 4.1.5) requires Deputy Heads to approve a three-year departmental security plan that “sets out strategies for meeting departmental security requirements reflective of and contributing to government-wide security priorities,” and addresses security controls. A security control is defined as a “legal, administrative, operational or technical measure for satisfying security requirements.” Appendix A of the PGS lists IT security and business continuity management (BCM) among the security controls to be addressed in departmental security plans. These controls help assure the security of IT systems (of which ADSs are an example) throughout their lifecycle and the continuity of the critical services and activities they support (e.g., in the event of a disaster or other disruptive event).

The DSM details mandatory procedures for the security controls identified in the PGS and holds the Chief Security Officer and other designated officials responsible for ensuring they are incorporated into departmental security plans, implemented, and monitored. Of particular relevance to subsection 6.3.6 of the Directive are the procedures requiring the establishment of mechanisms supporting IT continuity management (subsection B.2.3.10) as well as business continuity plans, measures, and arrangements (subsection D.2.2.3).

While the measures required under the contingency requirement are well established in the PGS and DSM, the term “contingency” is not defined or described in these instruments. The Directive also does

not provide a definition. Framing the requirement in terms of IT and business continuity management, and making clear links to the PGS and supporting policy instruments, could facilitate interpretation, improve coordination with departmental security officials, and minimize duplication of compliance efforts. By moving away from positioning contingency planning as a unique requirement rather than one with clear anchors in other policy instruments, this can also contribute to policy coherence.

Timing of AIA release

The Directive requires federal institutions to complete and publish an AIA to the Open Government Portal. However, subsections 6.1.1 and 6.1.4 do not specify when AIAs must be published. While TBS has encouraged federal institutions to publish their AIAs prior to the production of a system, a timing for release is not explicitly set in policy. This creates uncertainty as to the appropriate timing of publication and risks weakening the Directive's transparency measures by permitting institutions to delay AIA release well into a system's lifecycle. In fact, some AIAs published to the Open Government Portal were not released prior to system production.

Deploying and using an automated decision system in the absence of a publicly available AIA can have negative consequences for public trust in AI use in the federal public sector. All clients subject to automated decision-making should have access to a completed AIA before a system is activated to make or influence a decision impacting them. The earlier an AIA is released in the lifecycle of a system, the better for transparency and accountability.

This issue can be addressed by adjusting the requirement on AIA completion to clearly mandate publication prior to the production of a system. This would place current TBS guidance to federal institutions on a more formal footing and help ensure that clients and public stakeholders have timely access to information about automation projects in the government.

Policy Recommendations

In response to the challenges and risks discussed in the previous section, the following policy recommendations are proposed to guide the evolution of the Directive. The recommendations address the governance, objectives, scope, and requirements of the Directive, and propose additions to the AIA tool. For each recommendation, provisional amendments to the text of the Directive and AIA are proposed in Annex A.

Effective Date

- Periodic review: Replace the 6-month review interval with a biennial review and designate a role for the CIO of Canada to request a review based on need.

Scope

- Scope of policy: Expand the scope to cover internal services.
- Scope of policy: Clarify that the scope includes systems supporting administrative decisions.

Objectives and Expected Results

- Clients impacted by automated decision systems: Replace references to Canadians with more encompassing language such as clients.

Requirements

- Data governance: Introduce measures supporting the tracing, protection, and appropriate retention and disposition of data used and generated by an automated decision system.
- Model bias: Expand the pre-production testing requirement to cover model bias testing.
- Explanation: Establish explanation criteria in support of the explanation requirement.
- Peer review: Mandate the publication of a plain language summary of peer reviews and require their completion prior to system production.
- Contingency planning: Align the contingency requirement with relevant terminology established in Treasury Board security policy.
- Timing of AIA release: Mandate the release of AIAs prior to system production.

AIA Tool

- Reasons for automation: Expand the AIA to include questions prompting departments to articulate reasons for choosing to pursue automation.

Annex A: Proposed Amendments

Section 1 (Effective Date)

- Amendment to section 1.2: “This Directive will ~~have an automatic review process planned every 6 months after the date it comes into effect~~ be reviewed every two years, and as determined by the Chief Information Officer of Canada.”

Section 4 (Objectives and Expected Results)

- Amendment to section 4.1: “The objective of this Directive is to ensure that Automated Decision Systems are deployed in a manner that reduces risks to ~~clients Canadians~~ and federal institutions, and leads to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law.”

Section 5 (Scope)

- Amendment to section 5.1: “This Directive applies ~~only~~ to systems that provide external or internal services, in accordance with ~~as defined in~~ the Policy on Service and Digital.”¹⁰
- Amendment to section 5.2: “This Directive applies to any system, tool, or statistical models used to ~~support recommend~~ or make an administrative decision about a client.”

¹⁰ While the scope of the TB Policy on Service and Digital covers all internal services, it only defines “internal enterprise services”, a subset of internal services. So it’s more appropriate to frame the updated section as being “in accordance with” the Policy. Future updates to the Policy could address this gap by adding a definition of internal services which mirrors that of external services, “where the intended client is external to the Government of Canada” (Appendix A).

Section 6 (Requirements)

- Amendment to subsection 6.1.1: "Completing **and releasing** an Algorithmic Impact Assessment prior to the production of any Automated Decision System."
- Addition of new subsection under section 6.3 titled "**Data Governance**": "**Establishing measures to ensure that data used and generated by the Automated Decision System are traceable, protected, and appropriately retained and disposed of in accordance with the Directive on Service and Digital, Directive on Privacy Practices, and Directive on Security Management.**"
- Amendment to subsection 6.3.1: "Before launching into production, developing processes so that the data and information used by the Automated Decision Systems, **as well as the systems' underlying models**, are tested for unintended ~~data~~-biases and other factors that may unfairly impact the outcomes."
- Amendment to subsection 6.3.4: "Consulting the appropriate qualified experts to review the Automated Decision System **and publishing a plain language summary of the findings prior to the system's production**, as prescribed in Appendix C."
- Amendment to subsection 6.3.6: "Establishing ~~contingency strategies, plans, systems~~ **and/or measures processes to support IT and business continuity management**, as per Appendix C, **in accordance with the Directive on Security Management.**"
- Amendment to the title of subsection 6.3.6: "~~Contingency~~ **IT and Business Continuity Management**".

Appendix C (Impact Level Requirements)

- Amendment to the explanation measures for all four impact levels: "**This involves providing information describing:**
 - **The role of the system in the decision-making process;**
 - **The training and client data, their source and method of collection, if applicable;**
 - **The criteria used to evaluate client data and the operations applied to process it; and**
 - **The output produced by the system, and any relevant information needed to interpret it in the context of the administrative decision.**"
- Amendment to the peer review measures for impact levels II-III:
 - "**Consult** at least one of **the following experts and publish a plain language summary of the findings:**"
 - "**OR Publishing** specifications of the Automated Decision System in a peer-reviewed journal. **Where access to the published review is restricted, ensure that a plain language summary of the findings is openly available.**" (The latter entry would be positioned at the end of the list of options.)
- Amendment to the peer review measures for impact level IV:
 - "**Consult** at least two of **the following experts and publish a plain language summary of the findings:**"
 - "**Publishing** specifications of the Automated Decision System in a peer-reviewed journal. **Where access to the published review is restricted, ensure that a plain language summary of the findings is openly available.**"
- Amendment to the contingency planning measures for impact levels III-IV: "Ensure that **system recovery strategies, business continuity contingency plans, and/or other relevant security**

~~controls backup systems~~ are established in coordination with designated officials available should the Automated Decision System be unavailable.”

- Amendment to the title of the contingency planning section: “~~Contingency~~ IT and Business Continuity Management”.

Algorithmic Impact Assessment (AIA) Tool

- Addition of new series of questions on reasons for automation to section 2 of the AIA (Business Driver / Positive Impact):
 - What user need will the system address? [Free text]
 - How will the system be used to meet user needs? [Free text]
 - How effective will the system be in meeting user needs? [Slightly effective; Moderately effective; Very effective]
 - Please explain why you expect the system to achieve the level of effectiveness identified above. [Free text]
 - Please describe how you will ensure that the system is confined to addressing the user need identified above? [Free text]
 - Have alternative manual processes been considered? [Yes/No]
 - If manual processes were considered, why was automation identified as the preferred option? [Free text]
 - What is the consequence of not deploying the system? (Select all that apply) [Service cannot be delivered at all; Service cannot be delivered in a timely or efficient manner; Service costs are too high; Service quality is not as high; Service delivery cannot achieve performance targets; Other [Free text]]
- Amendment to the title of the Business Driver / Positive Impact section (page 2): “~~Business Driver~~ Reasons for Automation Positive Impact”.

References

A Local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools.

<https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9>

All-Party Parliamentary Group on the Future of Work, Report into the New Frontier: Artificial Intelligence at Work. <https://www.futureworkappg.org.uk/our-work/report-into-the-new-frontier-artificial-intelligence-at-work>

Artificial Intelligence Act. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>

Australian Government, Australia's AI Ethics Principles. <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>

Department of National Defence, EX-01 Report – Visible Minorities Recruitment Campaign. <https://www.canada.ca/en/department-national-defence/corporate/reports-publications/ex-01-report-visible-minorities-recruitment-campaign.html>

Edwards, Lilian and Veale, Michael, Enslaving the Algorithm: From a ‘Right to an Explanation’ to a ‘Right to Better Decisions’? (2018). IEEE Security & Privacy (2018) 16(3), pp. 46-54, DOI: 10.1109/MSP.2018.2701152, Available at SSRN: <https://ssrn.com/abstract=3052831> or <http://dx.doi.org/10.2139/ssrn.3052831>

General Data Protection Regulation (GDPR). www.gdpr-info.eu

Government of Canada, Open Government Portal, Algorithmic Impact Assessment. https://search.open.canada.ca/en/od/?search_text=algorithmic+impact+assessment

Immigration, Refugees and Citizenship Canada, Procedural fairness. <https://www.canada.ca/en/immigration-refugees-citizenship/corporate/publications-manuals/operational-bulletins-manuals/service-delivery/procedural-fairness.html>

Institute for the Future of Work (IFOW), Artificial Intelligence in Hiring: Assessing Impacts on Equality. <https://www.ifow.org/publications/artificial-intelligence-in-hiring-assessing-impacts-on-equality>

Institute for the Future of Work (IFOW), Mind the Gap: The Final Report of the Equality Task Force. <https://www.ifow.org/publications/mind-the-gap-the-final-report-of-the-equality-task-force>

Karanicolas, Michael, To Err is Human, to Audit Divine: A Critical Assessment of Canada’s AI Directive (April 20, 2019). Journal of Parliamentary and Political Law, Volume XIV:1, 2019, Available at SSRN: <https://ssrn.com/abstract=3582143> or <http://dx.doi.org/10.2139/ssrn.3582143>

Loi pour une République numérique. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>

Office of the Privacy Commissioner of Canada, OPC’s Guide to the Privacy Impact Assessment Process. https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/gd_exp_202003/

Office of the Privacy Commissioner of Canada, Supporting public health, building public trust: Privacy principles for contact tracing and similar apps. https://www.priv.gc.ca/en/opc-news/speeches/2020/s-d_20200507/

Organisation for Economic Development and Cooperation (OECD), Artificial Intelligence (AI) Principles. <https://oecd.ai/en/ai-principles>

Privacy Act. <https://laws-lois.justice.gc.ca/eng/acts/P-21/>

Public Services and Procurement Canada, Invitation to Quality (ITQ) on a source list of suppliers to provide Canada with responsible and effective Artificial Intelligence (AI) services, solutions and products. https://buyandsell.gc.ca/cds/public/2022/01/10/cf980aa59a6aab594f6db6f47a49f667/ABES.PROD.PW_EE.B017.E34526.EBSU002.PDF

Scassa, Teresa, Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada’s Directive on Automated Decision-Making (October 30, 2020). Forthcoming, (2021) 54:1 University of British Columbia Law Review, Available at SSRN: <https://ssrn.com/abstract=3722192> or <http://dx.doi.org/10.2139/ssrn.3722192>

Shared Services Canada, May 2021 Briefing to the Office of the Minister of Digital Government.

[Statistics Canada, Principles of Necessity and Proportionality.](https://www.statcan.gc.ca/en/trust/address)

<https://www.statcan.gc.ca/en/trust/address>

Statistics Canada, Responsible use of automated decision systems in the federal government.

<https://www.statcan.gc.ca/en/data-science/network/automated-systems>

The Global Partnership on Artificial Intelligence (GPAI), Working Group on the Responsible Development, Use and Governance of AI, A Framework Paper for GPAI's Work on Data Governance.

<https://gpai.ai/projects/data-governance/gpai-data-governance-work-framework-paper.pdf>

The Globe and Mail, National Defence skirted federal rules in using artificial intelligence, privacy commissioner says. <https://www.theglobeandmail.com/canada/article-national-defence-skirted-federal-rules-in-using-artificial/>

Treasury Board Directive on Automated Decision-Making. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>

Treasury Board Directive on Open Government. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108>

Treasury Board Directive on Privacy Practices. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18309>

Treasury Board Directive on Security Management. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32611>

Treasury Board Directive on Service and Digital. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32601>

Treasury Board Foundation Framework for Treasury Board Policies. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=13616>

Treasury Board Framework for the Management of Compliance. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=17151>

Treasury Board Guideline on Service and Digital.

<https://www.canada.ca/en/government/system/digital-government/guideline-service-digital.html>

Treasury Board of Canada Secretariat, Algorithmic Impact Assessment Tool.

<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>

Treasury Board Policy on Government Security. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=16578>

Treasury Board Policy on Privacy Protection. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12510>

Treasury Board Policy on Service and Digital. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32603>

U.S. Equal Employment Opportunity Commission, EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness. <https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness>