Treasury Board of Canada Secretariat | Secrétariat du Conseil du Trésor du Canada

Canada

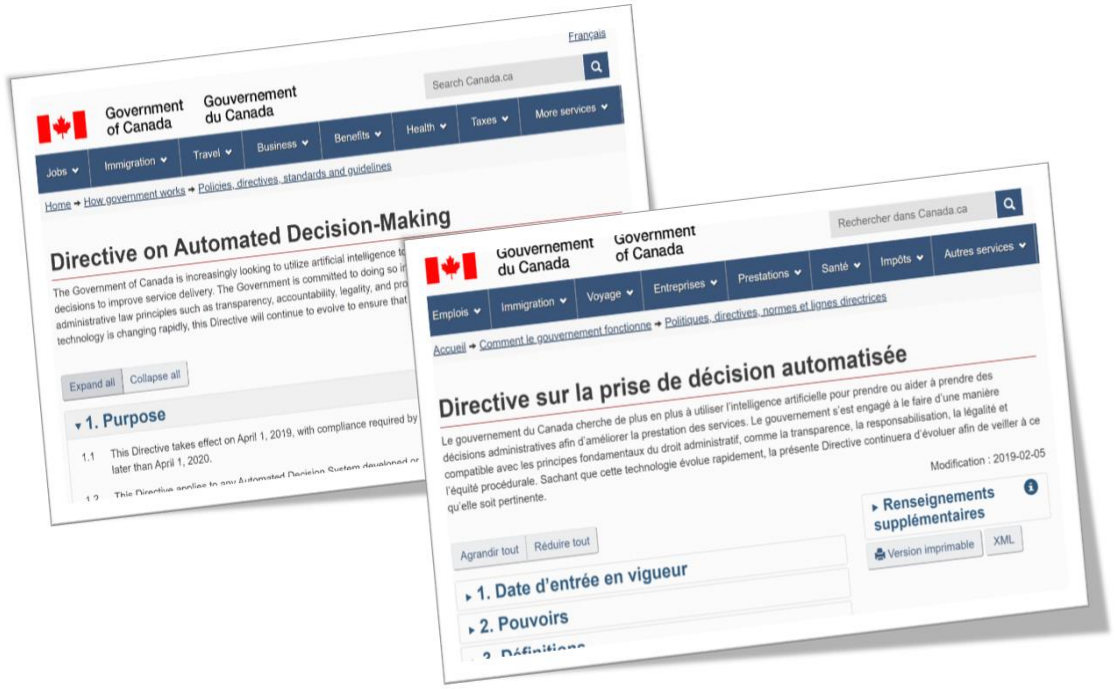
3rd review of the Directive on Automated Decision-Making: Summary of key issues, policy recommendations, and proposed amendments

Consultation deck
Fall 2022

Purpose

- **Provide an overview** of the 3rd review of the Directive on Automated Decision-Making (DADM).

- **Seek feedback** on policy recommendations and provisional amendments.



TBS is obliged to review the directive on a regular basis to ensure that it remains relevant and responsive to the evolving automation landscape in the federal government.

Context

- In recent years, governments and international organizations have pursued various regulatory efforts to manage the risks of artificial intelligence (AI) systems and foster algorithmic transparency and accountability.

- The Government of Canada (GC)'s approach to responsible AI promotes fairness and inclusion in automated decision-making by ensuring that the outputs of automated decision systems are explainable and unbiased.

- In 2019-20, over 300 AI projects were documented across 80% of federal institutions. Some use-cases automate decisions impacting service recipients within and outside the government.

Europe Is in Danger of Using the Wrong Definition of AI

Some intelligent systems are at risk of being excluded from oversight in the EU's proposed legislation. This is bad for both businesses and citizens.

Federal rules on AI too narrow and risk 'damaging public trust': Internal review

White House science advisers call for an "AI Bill of Rights"

Canada's opportunity to ensure AI remains a force for good

Research shows AI is often biased. Here's how to make algorithms work for all of us

LCO report comparing European Union, Canadian AI regulation stresses development of 'trustworthy AI'

The Directive on Automated Decision-Making

- The federal government sometimes uses computer systems to automate decision-making. In other cases, computer systems complete a portion of the analysis leading to a decision.

- When this involves service decisions that impact people's legal rights, interests, or privileges, requirements from the Directive on Automated Decision-Making apply.

- The requirements of the directive ensure that:

  - ✓ **People are informed about when and how automation is used;**
  - ✓ **People are provided with meaningful explanations about decisions affecting them;**
  - ✓ **The decisions are fair and accurate; and**
  - ✓ **The potential negative impacts of automation are identified and minimized.**

A Treasury Board directive sets mandatory requirements for how federal government organizations must operate.

Approach to the 3rd review

- The current review takes stock of the current state of the directive and identifies risks and challenges to the government's commitment to responsible AI in the federal public sector.

- The review examined the expanding range of services undergoing automation, identifying critical gaps and 'blind spots' that limit the directive's relevance and effectiveness in supporting transparency, accountability, and fairness in automated decision-making.

- Issues concerning terminology, feasibility, and coherence with other Treasury Board policy instruments have also informed the focus and direction of the review.

- Periodic reviews are not intended to be exhaustive. They seek to adapt the directive to pertinent trends in the Canadian and global AI landscape, while gradually refining the text of the instrument to support interpretation and facilitate compliance.

Overview of key issues identified in the 3rd review*

Scope

- External focus excludes automated decisions impacting federal employees.
- Language framing the scope requires clarification.

Periodic review

- Current 6-month timeframe for review creates policy and operational challenges.

Clients impacted

- Reference to Canadians in some parts of the directive does not recognize other potential clients.

Data governance

- Quality assurance measures do not address the need to trace, protect, and retain and dispose of data used and generated by a system.

Model bias

- Bias testing measures are limited to data – they do not account for other possible sources of bias such as the model.

Inclusion

- Measures supporting intersectional approaches to the design and implementation of systems are lacking.

Explanation

- Criteria for what constitutes a meaningful explanation are absent.
- Approach to publishing explanations is unclear.

Reasons for automation

- A justification for the adoption of AI in relation to a program's needs and goals is currently not required.

Peer review

- Requirement to publish information about peer reviews is not included.
- Timing of peer review completion is unclear.

Contingency planning

- Terminology is misaligned with Treasury Board security policy.

Timing of AIA release

- AIA requirements do not specify a timing for AIA release.

***** A more detailed description of these issues is available in the Annex.

High-level policy recommendations*

Governance

Expand the scope to cover internal services. (1)

Replace the 6-month review interval with a biennial review and enable the Chief Information Officer (CIO) of Canada to request off-cycle reviews. (3)

Transparency and accountability

Establish explanation criteria in support of the explanation requirement and integrate them into the Algorithmic Impact Assessment (AIA). (8)

Expand the AIA to include questions concerning an institution's reasons for pursuing automation and potential impacts on persons with disabilities. (9)

Mandate the publication of complete or summarized peer reviews and require completion prior to system production. (10)

Mandate the release of AIAs prior to the production of a system. (12)

Quality assurance

Introduce measures supporting the tracing, protection, and appropriate retention and disposition of data used and generated by a system. (5)

Expand the bias testing requirement to cover models. (6)

Mandate the completion of Gender Based Analysis Plus (GBA+) during the development of a system. (7)

Policy coherence

Clarify that the scope includes systems which make assessments related to administrative decisions. (2)

Replace references to Canadians with more encompassing language such as clients and Canadian society. (4)

Align the contingency requirement with relevant terminology established in Treasury Board security policy. (11)

***** Recommendations are numbered to facilitate reference to associated amendments in upcoming slides.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 1 (Effective Date)	Amend section 1.1: "This Directive **was amended and** takes effect on **[release date], 2023** ~~**April 1, 2019**~~, with compliance required by no later than **[date 6 months following release date], 2023** ~~**April 1, 2020**~~." Add section: "**Departments operating Automated Decision Systems developed or procured prior to [release date], 2023 will have 12 months to comply with requirements in subsections 6.2.3, 6.3.1, 6.3.X [data governance], 6.3.X [GBA+], and 6.3.4 of the Directive."**	TBS recognizes the challenge of adapting to new policy requirements while planning or executing projects that would be subject to them. In response, a 6-month 'grace period' is proposed to provide departments with time to plan for compliance with the amended directive. For systems that are already in place on the release date, TBS proposes granting departments a full year to comply with new requirements in the directive. Introducing this period would enable departments to plan for the integration of new measures into existing automation systems. This could involve publishing previously completed peer reviews or implementing new data governance measures for input and output data. During this period, these systems would continue to be subject to the current requirements of the directive.
DADM, Section 5 (Scope)	#1 - Amend subsection 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."	The adoption of automated decision systems to enhance internal services (e.g., hiring) in multiple federal institutions highlights the need to expand the scope of the directive to minimize risks to the rights, interests, and privileges of federal employees. The proposed amendment accomplishes this by including systems providing internal services, which are within the scope of the Policy on Service and Digital. The suggestion to replace 'as defined in' with 'in accordance with' arises from the fact that the policy (Appendix A) only defines 'internal enterprise services', a subset of internal services. (Consult the Annex for an overview of GC internal services.)
DADM, Appendix A (Definitions)	#1 (new) - Add definition: "**Internal service: A service where the intended client is internal to the Government of Canada. This includes federal employees.**"	The directive does not define internal services, many of which could become subject to it under an expanded scope. The Policy on Service and Digital only defines external and internal *enterprise* services. The proposed definition draws on the definition of external services and emphasizes that clients internal to the federal government include federal employees. (There are other types of internal services such as internal enterprise services where the client is an institution rather than an individual employee.) Clarifying the meaning of this term would ensure consistent interpretation and application of the amended directive.
DADM, Appendix B (Impact Assessment Levels)	#1 (new) - Add a new impact area under each of the four impact levels (I-IV): "**the equality, dignity, and autonomy of federal public servants**".	The impact scheme in Appendix B of the directive does not acknowledge the potential impacts of automation on federal public servants in the workplace. The proposed update identifies equality, dignity, and autonomy as three critical conditions of work that departments should uphold when seeking to introduce automation into the workplace. The principles draw on the Good Work Charter of the Institute for the Future of Work (IFOW), which provides an organizing framework for aspiration, alignment, and action to shape a fairer future of better work. Adding this impact area to the existing list adapts the directive's approach to impact assessment to an expanded scope, which would include internal services impacting federal public servants.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
AIA (Consultations Section)	#1 (new) - Amend the question to add new options to the lists for internal and external stakeholders: "Will you be engaging with any of the following groups? Internal Stakeholders (Strategic Policy and Planning, Data Governance, Program Policy, etc.): "**Digital Policy, Human Resources, TBS Office of the Chief Human Resources Officer, TBS Office of the Chief Information Officer**" External Stakeholders (Civil Society, Academia, Industry, etc.): "**Bargaining Agents**, **Governments in Other Jurisdictions, International Organizations, Clients or their Representatives**"	TBS is seeking to identify new internal and external stakeholders to account for use-cases involving internal services (which may be subject to the amended directive) and highlight previously missing types of stakeholders. These changes will better equip departments to identify and engage potentially relevant stakeholders within their respective institutions and outside the government on their automation projects. The lists are not intended to be exhaustive. They provide examples of offices and institutions that departments should consider consulting early in the project/system lifecycle to ensure alignment with applicable laws and policies, and to identify best practices and lessons learned in other organizations.
AIA (About the Decision Section)	#1 (new) - Amend the question: "Does the decision pertain to any of the categories below (check all that apply):" by adding a new option to the list: "**Employment (recruitment, hiring, promotion, performance evaluation)**"	The proposed option identifies employment as a key category of use-cases that would fall within the purview of the amended directive, which would apply to automation in internal services involving administrative decision-making. Recruitment, hiring, promotion, and performance evaluation are all examples of such services. While there are other types of internal services that may become subject to the directive, the third review seeks to prioritize the assessment and mitigation of risks arising in employment contexts.
DADM, Section 5 (Scope)	#2 (updated) - Amend section 5.2: "This Directive applies to any system, tool, or statistical model~~s~~ used to ~~**recommend or**~~ make an administrative decision **or a related assessment** about a client."	The distinction between recommending and making a decision is not clear. The current framing has caused confusion as to the conditions under which an automated system involved in an administrative decision-making process would be subject to the directive. The term 'recommend' obscures the intent to include any automated system influencing a decision-making process within the scope of the instrument (this is reflected in the definition of automated decision system in Appendix A of the directive). It can be misinterpreted as setting a high applicability threshold for systems that aren't making a decision. The proposed amendment introduces the concept of assessment to establish a clear threshold for evaluating whether cases of partial automation fall within the scope of the directive. The various ways in which a system can make an assessment about a client will be laid out in guidance. It's worth noting that the amendment does not modify the current scope; it merely expresses it in more suitable terms.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 1 (Effective Date)	#3 - Amend subsection 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."**	A more flexible review mechanism would help address the policy and operational challenges of the current provision. A two-year period would better account for the lengthy amendment process of a directive, while alleviating the capacity burdens imposed by a 6-month review interval. The proposed approach would also give policymakers more time to gather evidence on AI adoption and compliance across the GC, while bringing more stability and predictability to federal institutions subject to the directive. By enabling the CIO of Canada to request reviews at any time, the amendment ensures that the directive can remain responsive to needs as they arise.
DADM, Section 4 (Objectives and Expected Results)	#4 (updated) - Amend 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 Canadian society**, **and** leads to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian law."	Administrative decisions within the scope of the directive do not just concern Canadian citizens. They can also impact permanent residents, asylum seekers, visa applicants, or other individuals (or businesses) receiving a service from the federal government. It would therefore be more appropriate to use an encompassing term like 'clients' in the objective statement and throughout the instrument. This language would also support the proposed expansion of the scope of the directive to include federal employees, who aren't exclusively composed of Canadian citizens. It would also support alignment with the terminology used in the Policy on Service and Digital: the definition of client in the policy accurately reflects the broad range of people – Canadian citizens and others – who may be subject to automated administrative decisions. TBS also proposes adding a reference to Canadian society to highlight the social dimensions of automated decisions, which could have systemic impacts on certain communities or on the Canadian public at large. Overall, these changes communicate the government's commitment to inclusion, reinforce democratic accountability, and ensure consistency with existing policy terminology.
DADM, Section 6 (Requirements)	#5 - Add new subsection under 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.**"	While it establishes requirements supporting quality management and bias mitigation for data used by automated decision systems, the directive does not address the broader need to govern both the inputs and outputs of systems throughout their lifecycle. In particular, there is a need to ensure that data used and generated by systems are traceable, protected, and retained and disposed of appropriately. Where such data includes personally identifiable information, or can lead to such information in combination with other data, the directive should help ensure they are not shared, reused, or otherwise handled without the requisite authorities. The proposed subsection would enshrine safeguards for these unique data, ensuring that they are traceable (e.g., for audits, reviews, litigation, explanation), protected (e.g., from unauthorized sharing or use), and retained and disposed of appropriately. These measures would advance a more holistic approach to the governance of AI data.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 6 (Requirements)	#6 - Amend 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."	Bias in AI can arise from multiple sources including the data used to train a system and the model used to process it. The directive requires pre-production testing of input data for bias, but it overlooks the possibility that bias can also result from the assumptions and parameters built into a model. Achieving the desired outcomes of the directive demands oversight not only of input (and output) data, but also of the model used to derive the outputs supporting or constituting decisions. The proposed amendment would ensure that model-related issues are addressed early in the lifecycle, prior to system deployment.
DADM, Section 6 (Requirements)	#7 (new) - Add new subsection under 6.3: "**Completing a Gender Based Analysis Plus during the development of the Automated Decision System, as prescribed in Appendix C.**"	The current directive does not explicitly require departments to undertake a GBA+ for their automation projects. However, the AIA asks users whether they have undertaken a GBA+ for data collected for or used by the automated decision system. As it only applies the GBA+ methodology to input data, the effectiveness of this question in ensuring equitable practices in the development and use of automated decision systems may be limited. Establishing a dedicated requirement for GBA+ would ensure consistency in the application of the GBA+ lens to automated decision-making, and help foster intersectional approaches to the design, development, and use of automated decision systems. The elements proposed under Appendix C would help ensure a broad application of the methodology to data, systems, decisions, and other elements of an automation project. TBS has proposed this measure for systems at impact level II-IV, considering that GBA+ results are unlikely to be significant for level I systems.
DADM, Appendix C (Impact Level Requirements)	#7 (new) - Add new section on Gender Based Analysis Plus. Requirements in this area would apply to systems at impact levels II-IV: "**Ensure that your Gender Based Analysis Plus addresses the following issues:** • **Impacts of the automation project (including the system, data, and decision) on gender or other identity factors;** • **Planned or existing measures to address risks identified through the Gender Based Analysis Plus.**"	See rationale for GBA+ amendments (section 6).

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Appendix C (Impact Level Requirements)	#8 (updated) - Amend the explanation measures for impact level I: "In addition to any applicable legal requirement, ensuring that a meaningful explanation is **published** ~~**provided**~~ for common decision results. **The explanation should provide a general description of:** ~~**This involves providing information describing**~~ • **The role of the system in the decision-making process;** • ~~**The training and client**~~ **Input data,** ~~**their**~~ **its source and method of collection**~~**, if applicable**~~**;** • **The criteria used to evaluate input** ~~**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.** This **information should be made available in plain language through the Algorithmic Impact Assessment** ~~**can include providing the explanation via a Frequently Asked Questions section**~~ **and discoverable via** ~~**on**~~ a **departmental** website."	The current explanation requirement does not specify what constitutes a meaningful explanation. It is amenable to many interpretations, which creates several problems for federal organizations, TBS policy leads, and clients. The lack of clarity as to the information required to meet the requirement could result in inconsistent practices, which could lead to incomplete explanations and disparities in the treatment of clients. This also creates an ad-hoc approach to explainability in the government, with federal organizations seeking interpretive guidance from TBS policy leads on a case-by-case basis. This is burdensome not only for TBS but also for programs seeking to ensure effective compliance with the requirement. Inspired by France's *Loi pour une République numérique*, the proposed amendment formulates explanation criteria designed to address the need to safeguard a client's right to a fair and impartial decision-maker, and to reasons for decisions impacting them. The criteria would position the directive to better account for the digital character of automated decisions, which demands unique measures to ensure institutions looking to augment or replace human decision-makers can continue to meet the standards of administrative law. The criteria for impact level I are designed to reflect expectations for public explanations. The proposed text clarifies that explanations for level I systems should be published in plain language as part of the AIA and made discoverable on a departmental website. This is intended to strengthen the discoverability and accessibility of explanations of common decision results. As with any information in the AIA, public explanations would not include any personal or sensitive information. They also do not replace explanations addressed to a client, which departments must still provide to meet legal requirements and as required for impact levels II-IV.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Appendix C (Impact Level Requirements)	#8 continued (updated) - Amend the explanation measures for impact levels II-IV: ”In addition to any applicable legal requirement, ensuring that a meaningful explanation is provided **to the client** with any decision that resulted in the denial of a benefit~~, a~~ **or** service, or **involved a** ~~**other**~~ regulatory action. **The explanation should inform the client in plain language of:** • **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; and** • **A justification of the administrative decision.** **A general description of these elements should also be made available through the Algorithmic Impact Assessment and discoverable via a departmental website.**”	The proposed explanation criteria for impact levels II-IV help refine the distinction between public and private explanations. Explanations for systems at these levels should be addressed to clients in order to enable them to understand and contest a decision, and to meet procedural legal obligations. These explanations typically concern a specific individual – the subject of an automated decision – and are therefore likely to implicate procedural rights. The criteria mainly differ from those proposed for impact level I in that they would require departments to articulate the reasoning behind a decision. This criterion is intended to ensure clients understand why, given the outputs of a system (and potentially the judgement of an officer), a decision was decided a certain way. Since such justifications always pertain to a specific individual, this criterion is not being proposed for level I requirements, which mainly support public explanations. Drawing on the language proposed for level I explanations, TBS is also seeking to ensure that a public version of explanations provided to clients is drafted and published as part of the AIA. Providing public stakeholders with a "general description" of the same elements laid out in the criteria would strengthen algorithmic transparency and accountability. It would also enhance the effectiveness of the AIA itself, which would ask users to articulate how their systems are making or supporting decisions to a public audience, further expanding opportunities for public dialogue on the use of AI in the federal government.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
AIA (About the Data Section)	#8 (new) - Add the following question: "**Please describe the input data collected and used by the system, its source, and method of collection. [Free text]**"	This question is drawn from the proposed explanation criteria, which TBS is seeking to integrate into the AIA. See the rationales for updates to explanation requirements (Appendix C).
AIA (Business Driver / Positive Impact Section)	#9 (updated) - Add new series of questions on reasons for automation: • "**What user needs will the system address and how will this system meet them? If possible, describe how user needs have been identified. [Free text]**" • "**How effective will the system be in meeting user needs? [Slightly effective; Moderately effective; Very effective]**" • "**Please describe any improvements, benefits, or advantages you expect from using an automated system. This could include relevant program indicators and performance targets. [Free text]**" • "**Please describe how you will ensure that the system is confined to addressing the user needs identified above? [Free text]**" • "**Have alternative non-automated processes been considered? [Yes/No]**" • "**If non-automated processes were considered, why was automation identified as the preferred option? [Free text]**" • "**What ~~is~~ would be 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]]**" • "**Please describe any public benefits the system is expected to have. [Free text]**"	One of the guiding principles for responsible AI in the GC emphasizes the value of "starting with a clear user need and public benefit". The directive, however, does not ask departments to explain why they have chosen to introduce automation into a decision-making process. While the AIA includes questions asking departments to describe their automation project and identify relevant business drivers, departments are not expected to provide reasons justifying the necessity of automation for meeting specific user needs or providing public benefits. Similarly, the directive and AIA do not account for whether the use of a system will be appropriate for user needs and program objectives. The proposed series of questions positions the AIA as a space where departments can provide a rationale for their automation project, describing not only why it's necessary but also the degree to which it is compatible with user needs and 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 appropriate limits of AI use in administrative decision-making.
AIA (Risk Profile Section)	#9 (new) - Add new question: "**Will the use of the system pose significant risks for persons with disabilities? [Yes/No] If yes, please describe the accessibility risks and any planned or existing mitigation measures. [Free Text]**"	The proposed question integrates accessibility considerations into the AIA by ensuring that departments identify, assess, and mitigate the risks of automation to persons with disabilities. This allows departments to meet their legal obligations under the *Accessible Canada Act* and complements the proposed GBA+ measures to further strengthen fairness and inclusion in automated decision-making.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 6 (Requirements)	#10 (updated) - Amend subsection 6.3.4: “Consulting the appropriate qualified expert**s** to review the Automated Decision System **and publishing the complete review or a plain language summary of the findings prior to the system's production**, as prescribed in Appendix C.”	The absence of a mechanism mandating the release of peer reviews (or related information) creates a missed opportunity for bolstering public trust in the use of automated decision systems through an externally sourced expert assessment. Releasing at least a summary of completed peer reviews (given the challenges of exposing sensitive program data, trade secrets, or information about proprietary systems) can strengthen transparency and accountability by enabling stakeholders to validate the information in AIAs. The current requirement is also silent on the timing of peer reviews, creating uncertainty for both departments 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. The proposed amendments address these issues by expanding the requirement to mandate publication and specify a timing for reviews. Published peer reviews (or summaries of reviews) would complement documentation on the results of audits or other reviews that the directive requires project leads to disclose as part of the notice requirement (see Appendix C of the directive).
DADM, Appendix C (Impact Level Requirements)	#10 (updated) - Amend the peer review measures for impact levels II-III: "**Consult** at least one of **the following experts and publish the complete review or a plain language summary of the findings on a Government of Canada website**:"; "**OR** Publish~~**ing**~~ 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.) Amend the peer review measures for impact level IV: "**Consult** at least two of **the following experts and publish the complete review or a plain language summary of the findings on a Government of Canada website**:"; "Publish~~**ing**~~ 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 proposed amendments are intended to harmonize the peer review measures in Appendix C with the updated peer review requirement (subsection 6.3.4). They emphasize the need to publish a complete or summarized peer review on a Government of Canada website. For impact levels II-III, the option to publish specifications of the automated decision system in a peer reviewed journal has been positioned as an alternative to consulting one or more of the expert groups. In alignment with the approach taken under impact level IV, this equates journal peer review with the review conducted by any one of the qualified experts listed.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 6 (Requirements)	#11 - Amend 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.**” Amend the title of subsection 6.3.6 by replacing “Contingency” with “**IT and Business Continuity Management**”.	The measures required under the contingency requirement are well established in the Policy on Government Security (PGS) and Directive on Security Management (DSM). The term “contingency”, however, 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. The proposed amendments to Appendix C seek to align to the language proposed for the requirements section. But they also provide additional detail, drawing on mandatory procedures for security controls in the DSM, particularly in the context of IT and business continuity management.
DADM, Appendix C (Impact Level Requirements)	#11 (continued) - Amend 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.” Amend the title of this section by replacing “Contingency Planning” with “**IT and Business Continuity Management**”.	See rationale for contingency amendments (section 6).
DADM, Section 6 (Requirements)	#12 (updated) - Amend subsection 6.1.1: "Completing **and releasing the final results of** an Algorithmic Impact Assessment prior to the production of any Automated Decision System."	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. 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 without delay. The earlier an AIA is released in the lifecycle of a system, the better for transparency and accountability. Some AIAs published to the Open Government Portal were not released prior to system production. The proposed amendment addresses this issue by clearly stating the need to release an AIA prior to the production of a system.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale

AIA (Impact Assessment Section)	#1, #2, #8 (new) - Replace the first two questions currently in this section of the AIA with:
	• "**Which of the following best describes the type of automation you are planning?**
	  • **Full automation (the system will make an administrative decision)**
	  • **Partial automation (the system will contribute to administrative decision-making by supporting an officer through assessments, recommendations, intermediate decisions, or other outputs)**
	Please describe the role of the system in the decision-making process. [Free text]"
	
	Add the following questions:
	• "**Please describe the criteria used to evaluate client data and the operations applied to process it. [Free text]**"
	• "**Please describe the output produced by the system and any relevant information needed to interpret it in the context of the administrative decision. [Free text]**"
	• "**Will the system perform an assessment or other operation that would not otherwise be completed by a human? [Yes/No] If yes: Please describe the relevant function(s) of the system. [Free text]**"
	• "**The impacts that the decision will have on the equality, dignity, and autonomy of federal public servants will likely be: [Little to no impact, moderate impact, high impact, very high impact] Please describe why the impacts resulting from the decision are (as per the option selected above) [Free text]**"
	
	Amend the question: "Will the system be **making** ~~**replacing**~~ ~~**human**~~ decisions **or assessments** that require judgement or discretion?"
		There is a need to address redundancies and contradictions in the AIA's questions on the role of a system in a decision-making process: the first two questions in this section are not only duplicative, but also falsely associate partial automation with a higher level of risk than full automation. The proposed updates address this concern in accordance with TBS's proposal to amend the language used to describe the scope of the directive. TBS proposes a consolidated question that identifies the type of automation being pursued and asks users to further describe the role of their system in decision-making. This description would form part of the public explanation required under Appendix C of the amended directive. See the rationales for amendments to explanation requirements under Appendix C.
		
		The updates to this section also introduce new questions evaluating whether a system would perform novel assessments which may not be feasible in a non-automated process, and any potential impacts on federal public servants (in alignment with the updates proposed under Appendix B). The question on novel assessments is intended to identify new processes that may not be part of an existing decision-making process. Such processes may carry unique risks and should therefore be identified as part of the AIA.
		
		The question on impacts on federal public servants is intended to account for the application of the amended directive to internal services impacting federal public servants. This corresponds to the impact area proposed under Appendix B of the directive. Consult the rationale for this amendment for more information.
		
		Like the request to articulate the role of the system in the decision-making process, the questions concerning the input and output data are drawn from the proposed explanation criteria. Here, too, consult the rationales for amendments to explanation requirements under Appendix C.

Proposed amendments to the directive

Instrument & section	Proposed amendment (provisional text)	Rationale
DADM, Section 10 (References)	#1, #7, #9 (new) - Amend section 10.1: "***Accessible Canada Act***" Amend section 10.2: "**Policy on People Management**"	The proposed additions to the References section are intended to complement three amendments: the expansion of the directive's scope to cover internal services, the requirement to complete a GBA+ during the development of an automated decision system, and the introduction of a question on accessibility impacts into the AIA. The Policy on People Management governs the organization and management of the federal public workforce. It establishes rules for the delivery of internal services in the human resources domain, many of which would fall within the purview of the amended directive (to the extent that they involve administrative decisions). These requirements would have to be factored into impact assessments for systems deployed in this domain (e.g., to support recruitment, hiring, or performance management). The policy is therefore relevant to the effective adoption of the measures of the amended directive. The *Accessible Canada Act* aims to realize a barrier-free Canada by 2040. The legislation benefits all Canadians, especially persons with disabilities, through the proactive identification, removal, and prevention of barriers to accessibility in a range of areas including Information and Communication Technologies (ICT). The proposed requirement to complete a GBA+, and the new AIA question concerning the potential impact of an automation project on persons with disabilities, draw on legal obligations established in the Act. The Act's relevance as a reference point in the directive goes beyond these measures, however. The Act prioritizes accessibility in the design and delivery of programs and services, thus making it a relevant source of law for any potential use of automated decision systems in service delivery.

Expected outcomes

Implementing the proposed amendments would:

- Ensure automated decision systems affecting civil servants are fair and inclusive.
- Reinforce transparency and accountability.
- Strengthen protections against discrimination and harm.
- Clarify requirements and support operational needs.

Discussion questions

- Are there any critical issues that the current review does not take into consideration?
- Are the proposed amendments to the directive and AIA clear and appropriately justified?
- Do you foresee any problems with amending the directive and AIA as proposed?
- Are there any key federal or external stakeholders that TBS should engage as part of this process?
- What issues should TBS consider prioritizing in the next review of the directive?

Next steps

- Working in the open, collaborate with national and international stakeholders to identify and address issues ahead of the policy amendment process.

March	Spring-Fall	Fall 2022-Winter 2023
Stage 1: Preliminary OCIO consultation	**Stage 2: Engagement with stakeholders**	**Stage 3: Policy amendments**
• Ensure alignment with privacy, security, open government, and digital policy • Raise awareness of 3rd review (ongoing)	• Refine recommendations and amendments • Consult departments, Office of the Privacy Commissioner of Canada, service and human resources officials, and bargaining agents • Engage with researchers, civil society, and other governments	• Start OCIO gated policy stewardship process • Seek senior committee endorsements • Seek GC CIO and Secretary approval • Publish amended directive • Support departments with compliance (ongoing)

Contact

Treasury Board of Canada Secretariat
Office of the Chief Information Officer
ai-ia@tbs-sct.gc.ca

Annex

Key issues identified in the 3rd review

- **Scope:** The scope of the directive excludes automated administrative decisions impacting federal employees. This creates vulnerabilities for employees subject to automation in hiring, performance evaluation, or other decisions supporting internal service delivery. As well, the terms used to frame the scope of the directive have caused confusion as to the conditions that trigger it.

- **Periodic review:** The 6-month review interval presents policy and operational challenges to TBS. This is due to the length of consultation and approval processes; impact of regular reviews on team capacity; the relatively slow pace of automation adoption in the GC; and the uncertainties arising from frequent changes to administrative policy.

- **Clients impacted by automated decision systems:** Where it specifically addresses Canadians, the directive falls short of recognizing its potential applicability to cases impacting other clients in Canada or abroad (e.g., permanent residents, refugees, citizens of other countries).

- **Data governance:** While the directive includes provisions supporting the management of data collected for and used by a system (e.g., to minimize bias, assure quality), it does not establish measures supporting the traceability, protection, and appropriate retention and disposition of this data. This is also needed for system outputs (e.g., recommendations, scores), which are not addressed in the directive. Both types of data could pose privacy or security risks if shared, reused, retained, or disposed of inappropriately.

See the section on ‘Challenges & Risks' in the full report for additional information on each issue.

Key issues identified in the 3rd review

- **Model bias:** The quality assurance requirements of the directive do not address bias arising from the model underlying a system (rather than the data used to develop it). This could lead users to overlook pre-production model testing.

- **Inclusion:** While it requires bias testing, the directive does not prescribe any frameworks or methods for examining inclusion in automation. This limits the directive's potential to ensure fair and equitable decision-making.

- **Explanation:** The explanation requirement does not specify what constitutes a 'meaningful explanation'. The lack of explanation criteria could lead to inconsistent interpretation and application. The requirement also does not prescribe an approach to publication.

- **Reasons for automation:** The directive does not account for the purpose and scope of automation projects. This gap leaves clients and public stakeholders without a clear justification of a program's decision to adopt AI and description of how a system will be deployed to meet user needs and program goals.

- **Peer review:** The absence of a requirement to publish information about peer reviews constitutes a missed opportunity for the GC, which could leverage this mechanism to bolster public trust in automated decision systems in use within government. The appropriate timing of peer reviews is also unclear.

See the section on 'Challenges & Risks' in the full report for additional information on each issue.

Key issues identified in the 3rd review

- **Contingency planning:** The terminology used in this requirement is not aligned with what is well established in Treasury Board security policy, which sets security controls for IT, business continuity management, and other areas. This mismatch could lead to duplication of compliance efforts and negatively impact policy coherence.

- **Timing of AIA release:** The directive does not specify a timing for the release of AIAs. The requirement on AIA release only establishes the format and location of publication. This creates uncertainty as to the appropriate timing of release and risks weakening the directive's transparency measures by allowing institutions to delay AIA release well into a system's lifecycle. This has negative consequences for public trust in AI use in the federal public sector.

See the section on 'Challenges & Risks' in the full report for additional information on each issue.

GC guiding principles for responsible artificial intelligence

1. **Understand & measure** the impact of using AI
2. **Transparency** about how and when we are using AI
3. **Meaningful explanations** about AI in decision making
4. **Be as open as we can** by sharing source code, training data, and other relevant information
5. **Provide sufficient training** that enables public servants to develop and use AI solutions that have responsible design, function, and implementation

Overview of the Directive on Automated Decision-Making

Directive on Automated Decision-Making Requirements

Algorithmic Impact Assessment

Understand

- AIA before production
- AIA when scope changes
- Release of AIA results

Transparency

Communicate

- Notice before decision
- Explanation after decision
- Access to components
- Release of source code
- Documentation of decisions

Quality assurance

Prevent

- Testing and monitoring of outcomes
- Data quality
- Peer review
- Employee training
- Contingency
- Security
- Consultation with legal services
- Human intervention

Recourse

Correct

- Recourse options to challenge decisions

Overview of directive requirements

- Released in 2019, the directive seeks to ensure transparency, accountability, and procedural fairness in the use of automated decision systems in the federal government.
- The scope of the directive covers systems used to make or support administrative decisions impacting external clients (e.g., citizens, businesses). It applies to systems developed or procured as of April 1st, 2020.
- The directive formalizes algorithmic accountability by holding Assistant Deputy Ministers (ADMs) overseeing relevant automation projects responsible for complying with the policy's requirements.
- Federal institutions subject to the directive are required to complete and publish an Algorithmic Impact Assessment (AIA) to the Open Government Portal. The AIA tool is a questionnaire that determines the impact level of an automated decision system.

Overview of directive requirements

- The impacts of automating an administrative decision are classified into four levels, ranging from Level I (little impact) to Level IV (very high impact). The AIA helps identify risks and assess impacts in a broad range of areas related to the rights and interests of individuals and communities.
- The directive establishes quality assurance measures to help ensure the legality of an automation project, quality of 'input' data, system security, human oversight, peer review, and employee literacy.
- Systems in production must be monitored to guard against unintentional outcomes and ensure compliance with applicable policy and legislation.
- The directive requires federal institutions to provide clients subject to automated decision-making with an appropriate recourse mechanism enabling them to contest a decision.
- TBS uses multiple governance mechanisms to ensure compliance with the directive, including the Framework for the Management of Compliance, departmental concept cases, enterprise architecture proposals, and Treasury Board submissions.

Examples of system functions in a decision-making process

Recommendations

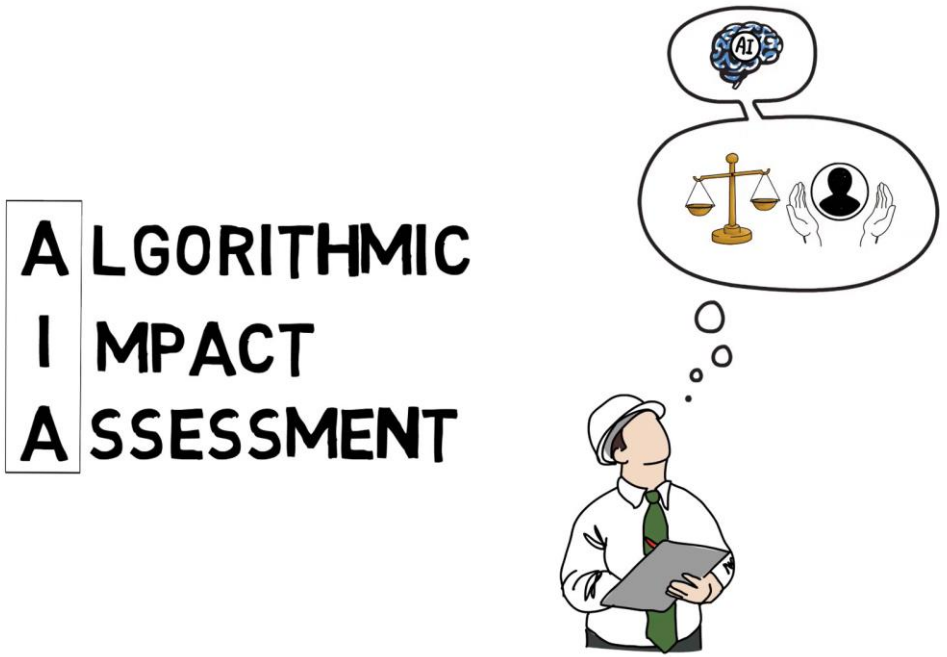
- present relevant information to the decision-maker;
- alert the decision-maker of unusual conditions;
- present information from other sources (“data matching”);
- provide assessments, for example by generating scores, predictions, or classifications;
- recommend one or multiple options to the decision-maker;
- make partial or intermediate decisions as part of a decision-making process; or
- make the final decision.

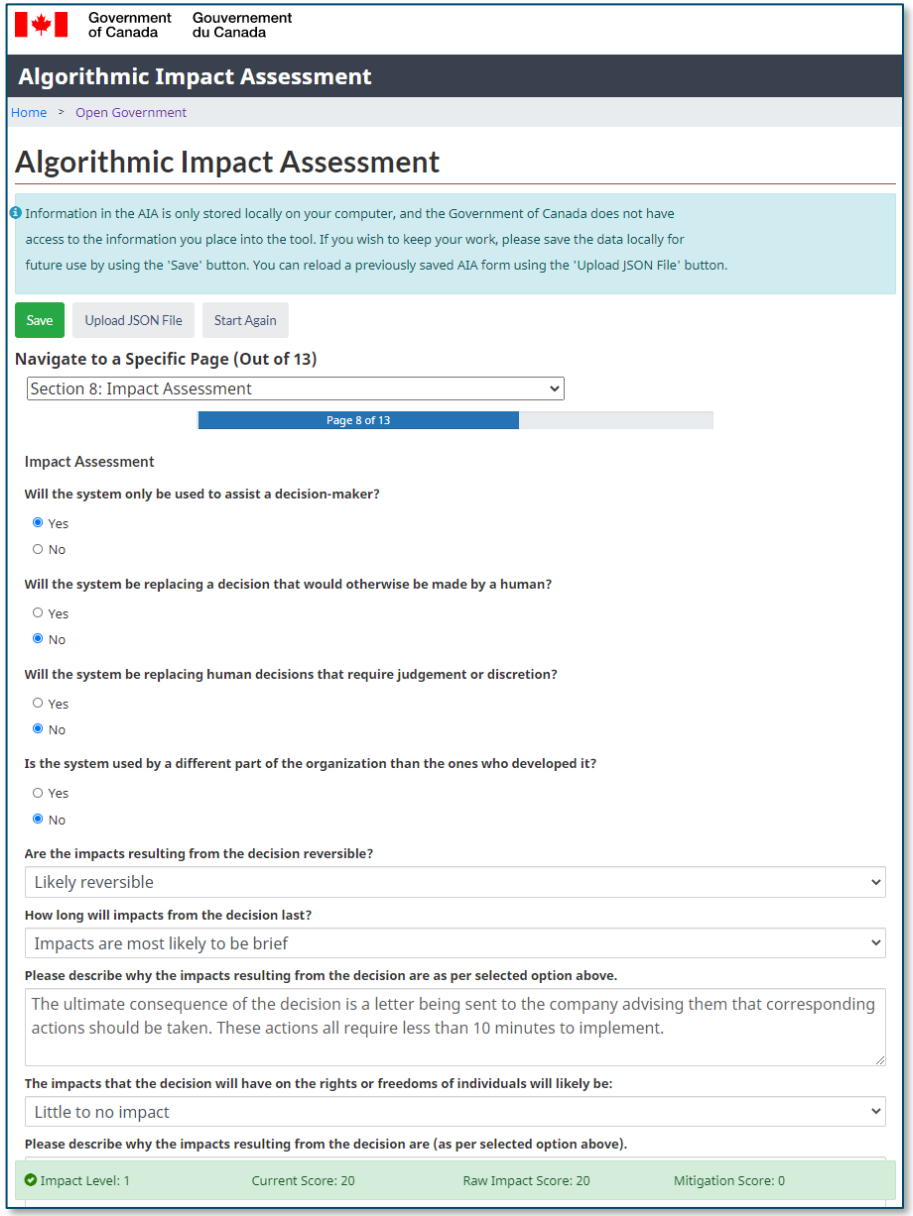
Decisions

Algorithmic Impact Assessment (AIA)

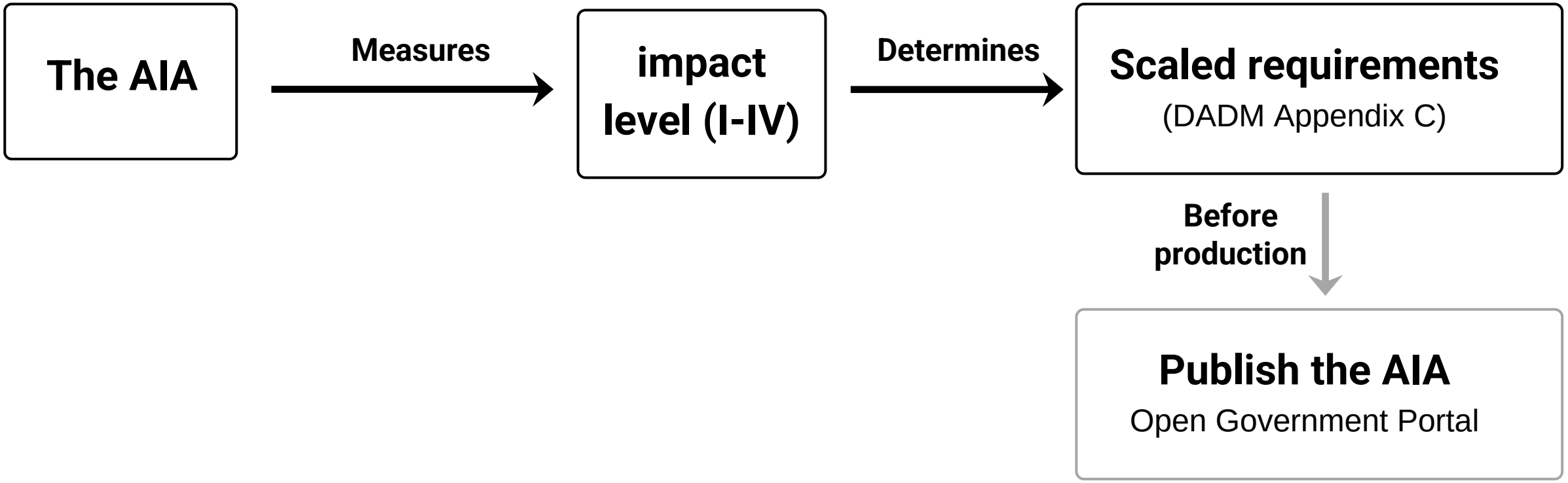
https://open.canada.ca/aia

Description, instructions and scoring methodology explained at:
https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html



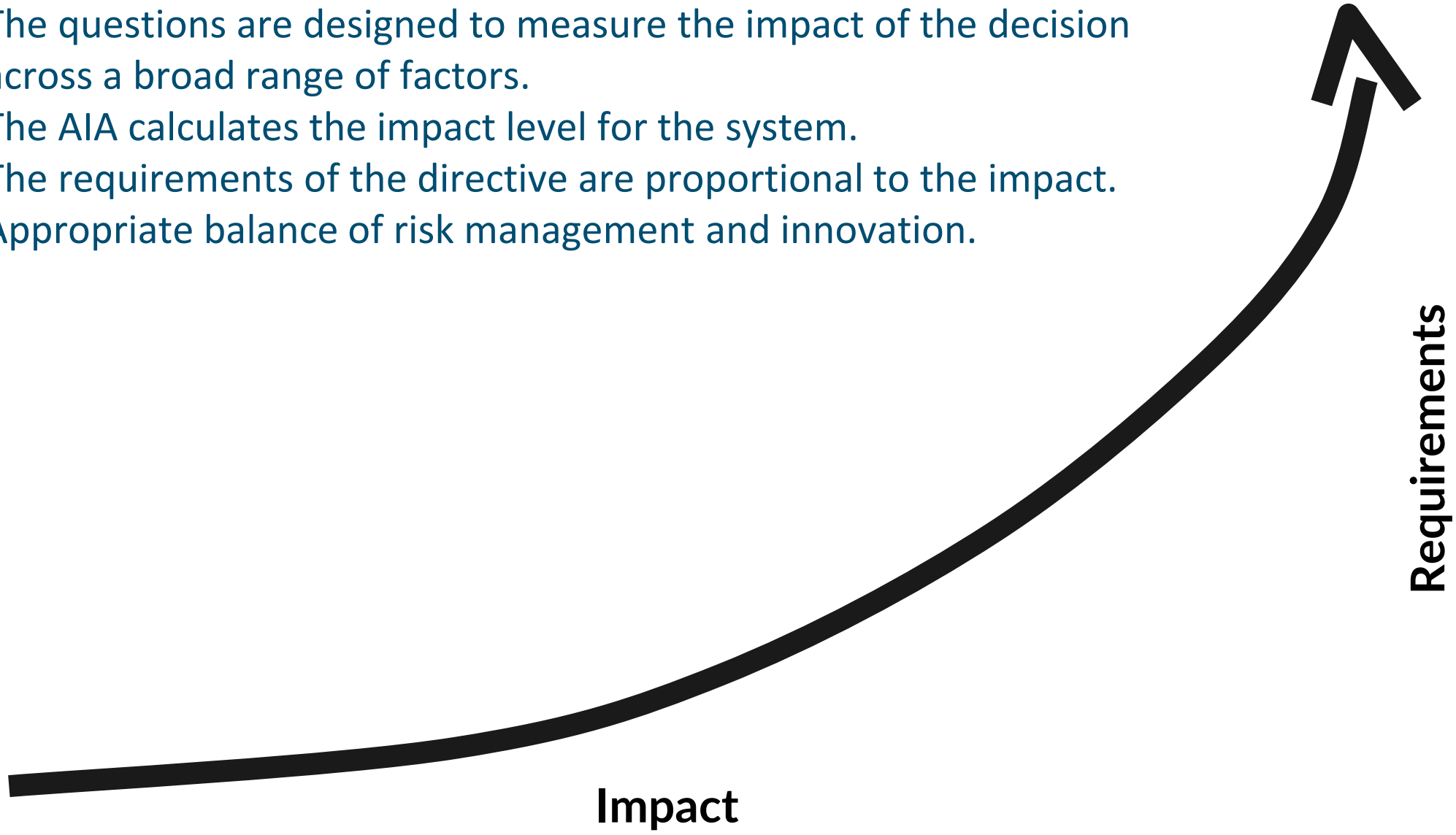


The AIA: overview of process



The AIA: progressive requirements

- The questions are designed to measure the impact of the decision across a broad range of factors.
- The AIA calculates the impact level for the system.
- The requirements of the directive are proportional to the impact.
- Appropriate balance of risk management and innovation.



The AIA: progressive requirements (example from Appendix C)

Requirement	Level I	Level II	Level III	Level IV
Human-in-the-loop	Decisions may be rendered without direct human involvement		Decisions cannot be made without having specific human intervention points during the decision-making process; and must be made by a human	
Notice	None	Plain language notice posted through all service delivery channels in use (Internet, in person, mail or telephone)	Plain language notice through all service delivery channels in use (Internet, in person, mail or telephone). In addition, publish documentation on relevant websites about the automated decision system, plain language, describe: • How the components work; • How it supports the administrative decision; • Results of any reviews or audits; and • A description of the training data, or a link to anonymized training data if this data is publicly available	

Internal services in the GC

- The scope of the Policy on Service and Digital covers all services, including those delivered to federal employees and others within the government.

- The Guideline on Service and Digital defines internal services 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.”

- The GC Service Inventory provides the authoritative list of GC services subject to the Policy on Service and Digital. TBS has identified a widespread need for consistent inclusion of internal services in the Inventory.

- The categories and terms used to classify and define internal services are established in the Canadian Government Reference Model (CGRM) and GC Strategic Reference Model (GSRM). The CGRM and GSRM identify internal programs (e.g., HR management) and associated internal services (e.g., recruitment) and service outputs (e.g., resources).

- The reference models and GC Service Inventory can be leveraged to inform TBS about the services which could be subject to the directive, should its scope be expanded to cover internal services.