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

Canada

4th review of the Directive on Automated Decision-Making

Overview and proposed modifications

Fall 2024

Purpose

- Provide an overview of the 4th review of the Directive on Automated Decision-Making

- Seek feedback on policy recommendations and proposed modifications

- Discuss proposed updates

Background

- Providing better programs and services for Canadians sometimes involves the use of artificial intelligence (AI) to support sound decision making

- The Government of Canada (GC) is committed to ensuring that the government's use of AI is governed with clear values, ethics and laws, and in accordance with human rights

- Treasury Board of Canada Secretariat (TBS) is responsible for providing government-wide direction on information and data governance, information technology, security, privacy and access to information.

- The Directive on Automated Decision-Making (directive) sets rules for how federal departments can use automated systems (including AI) to make or support decisions that impact the legal rights, privileges or interests of clients. For example:
  - determining eligibility for permits and benefits
  - assessing eligibility for entry to Canada
  - deciding to hire an individual to work in the public service
  - granting market authorization for a pharmaceutical product
  - launching an investigation into an individual's conduct

Directive on Automated Decision-Making

The directive seeks to ensure transparency, accountability and procedural fairness in the use of automated decision systems in the federal government.

It requires departments to:

- assess the impacts of automated decision systems

- be transparent

- ensure quality

- provide recourse on decisions

- report publicly on system effectiveness and efficiency

The directive came into effect in April 2019 and applies to systems developed or procured after April 2020.

Overview of the Directive on Automated Decision-Making

Directive on Automated Decision-Making Requirements

Algorithmic Impact Assessment	Transparency	Quality assurance	Recourse
Understand	*Communicate*	*Prevent*	*Correct*

Understand
- AIA before production
- AIA when scope or functionality changes
- Publication of AIA results

Communicate
- Notice before decisions
- Explanation after decisions
- Access to components
- Release of source code
- Documentation of decisions
- Publication of results in meeting program objectives
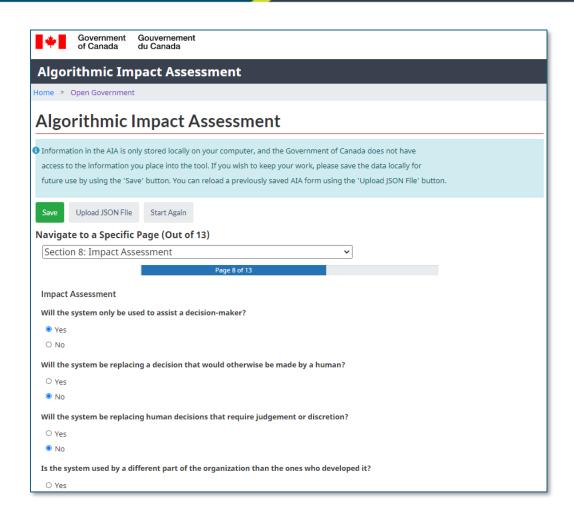
Prevent
- Testing and monitoring of outcomes
- Data quality
- Data governance
- Peer review
- GBA Plus
- Employee training
- Continuity management
- Security
- Consultation with legal services
- Human intervention

Correct
- Recourse options to challenge decisions

Algorithmic impact assessment (AIA)

- Mandatory risk assessment tool
- Determines the impact level of an automated decision-system
- 51 risk and 34 mitigation questions
- Scores based on many factors including systems design, algorithm, decision type, impact and data
- Developed based on best practices in consultation with both internal and external stakeholders
- Developed in the open, and available to the public for sharing and re-use under an open license

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

Reviewing the Directive

The directive must be reviewed every 2 years to keep the instrument relevant and reflect the evolving technology and regulatory landscape.

1st review (2020-21)	2nd review (2021-22)	3rd review (2022-23)	4th review (2024-25)
✓ Strengthen transparency and quality assurance ✓ Update references to policy instruments ✓ Clarify requirements	✓ Author guidelines supporting interpretation of requirements	✓ Expand scope ✓ Strengthen transparency and quality assurance ✓ Enable inclusive approaches ✓ Improve coherence with other policies ✓ Assess reasons for automation ✓ Assess impacts on persons with disabilities ✓ Clarify requirements	Underway summer 2024 to 2025 ❑ Support effective implementation ❑ Strengthen client protections ❑ Enhance assessment of impacts

4th review intended outcomes

Implementing the proposed updates would:

- Improve protections to clients and federal institutions

- Strengthen compliance with the directive

- Reinforce commitments to transparency and accountability

- Clarify and improve understanding of requirements and AIA questions to align with intent and interpretation in practice

- Reduce redundancies and misalignment across the TBS policy suite

Overview of key themes and issues

Three themes with 7 topics have been identified to address in the 4th review. Each theme has a goal and recommendations to achieve it.

Support effective implementation	Strengthen client protections	Enhance assessment of impacts
• Monitor policy implementation • Reduce the number of organizations excluded from directive • Adopt internationally recognized definition of AI	• Clarify obligations and enhance impact assessment of human rights • Strengthen protections and assessment of impacts for persons with disabilities • Identify banned uses	• Clarify and enhance the AIA

In addition to the changes across 3 themes, additional changes are proposed to improve clarity, reduce redundancies and align with other policy instruments. All proposed changes to the Directive and AIA are available on our GCwiki page

Support effective implementation

	Monitoring policy implementation	Excluded organizations	Definition of AI
Goal	Increase and verify departmental compliance with the directive to support improved outcomes for clients, federal institutions and Canadian society.	Increase the number of organizations that are subject to the directive to expand protections and reduce risks to clients, federal institutions, and Canadian society.	Align the GC definition of AI with a more recent and internationally-recognized definition, increasing compatibility and facilitating understanding.
Recommendations	• Add a new reporting requirement for departments to submit a report signed by the responsible assistant deputy minister (ADM) to TBS confirming compliance with the Directive. • Add clarity to the role of the CIO of Canada related to compliance monitoring. • Add a responsibility for TBS to publish an annual summary of the compliance reports on the Open Government Portal. • Add a requirement for the responsible ADM to approve the completed AIA prior to its publication.	• Remove the Excluded organizations subsection 9.1.1, such that the directive would apply to Agents of Parliament	• Remove the definition of AI from Appendix A of the directive • Modify the definition of AI in Appendix A of the Policy on Service and Digital to align with the OECD definition

Strengthen client protections

	Human rights	Persons with disabilities	Bans
Goal	Clarify obligations and enhance impact assessment of human rights	Strengthen protections and assessment of impacts for persons with disabilities, raising awareness of the impacts of automated decision systems	Identify explicit limits or define circumstances in which automated systems pose an unacceptable risk
Recommendations	• Modify the testing and monitoring requirements in the directive to more clearly reference human rights • Add targeted questions to the AIA to strengthen the consideration of impacts to a broader range of people	• Add a requirement to document system failures and take corrective actions • Add new AIA questions to capture compliance with existing accessibility standards and broaden the consideration of impacts	• Add a requirement in the Policy on Service and Digital to provide parameters for use • Add a requirement and an appendix to the Directive on Service and Digital and prepare supporting guidance to state uses of AI considered unacceptable by the GC

6. Examples of unacceptable AI uses

Unacceptable use	Example of unacceptable use case
To manipulate or deceive in a way that alters behaviour resulting in harm or impacts to individual autonomy and fundamental freedoms	• Using deepfake technology that can spread misinformation, manipulate public opinion, and lead to increased polarization.
To score or classify people in a way that leads to unjustified censorship or surveillance or that impacts freedom of expression, privacy, and autonomy	• Collecting social behaviour and inferred, predicted or observed personal characteristics over time to create a score that impacts individuals' and groups' ability to function in society.
Biometric categorization to infer personal information about individuals such as their race, political affiliation, religion, sexual orientation, gender identify and expression, and disability status	• Using physical characteristics such as hair and eye colour to infer a person's ethnicity.
Untargeted facial recognition scraping with the intent to create or expand databases	• Taking publicly available images from the internet for identity verification
Real-time and remote biometric identification systems used in public spaces, with limited exceptions when the risk of harm is outweighed by the benefit	• Using real-time facial recognition to identify individuals that are subject to a regulatory fine.
Emotion recognition	• Relying on AI to infer emotions with the goal of determining risk, capabilities or skills or to make a decision without human oversight.
Determining the risk of a person or group committing an offence based solely on AI profiling or AI assessment of personality traits	• Relying on the use of predictive policing techniques to forecast criminal activity.

Enhance assessment of impacts

AIA modifications	
Goal	Increase clarity and thoroughness of the AIA tool
Recommendations	• Add questions in AIA sections where gaps exist or that support other areas of the 4th review • Modify and add questions to respond to feedback and clarify intent • Editorial changes

All changes to the AIA can be found here

Tell us what you think

- We are seeking your input on the recommendations and directive edits as part of the 4th review of the directive.

- Please respond to the survey where you will be asked to:
  o confirm if the recommendations help to achieve the topic goals
  o share specific edits to the updated text of the directive and AIA
  o identify any concerns or gaps
  o provide input on the approach to the "bans" topic

- We recommend that you refer to the following documents as you complete the survey:
  o Overview of the 4th review of the Directive on Automated Decision-Making
  o Text changes to the Directive
  o Text changes to the AIA tool

- The survey will be open from November 19, 2024 to January 8, 2025

Next steps

- Review the 4th review materials and respond to the survey by January 8

Timeline

Fall 2024	Winter 2025	Spring 2025	Ongoing
Consultation			
Analysis & Incorporate Feedback			
	Governance & Approvals		
		Publish	
			Support Departments

Questions for discussion

- Do the proposed updates appropriately address the goals for the 7 identified topics? Why or why not?

- Are there any proposed changes to the Directive or AIA that are concerning?

- Is there a theme or topic that is missing?

Questions?

Please reach out to the TBS Responsible Data and AI team

(ai-ia@tbs-sct.gc.ca)