Digital Society (2025) 4:64
https://doi.org/10.1007/s44206-025-00221-7

ORIGINAL PAPER

Design versus reality: assessing the results and compliance of algorithmic impact assessments

Ana Brandusescu[1] · Renée E. Sieber[2]

Received: 17 October 2024 / Accepted: 9 July 2025

Abstract

Algorithmic impact assessments (AIAs) have become a dominant regulatory instrument in governing artificial intelligence (AI). While there are noteworthy examples across the global north, Canada's AIA is considered to be the best practice worldwide. When AIAs are studied, evaluations have been based on the assessment of the instrument and not the examination of their answers. We examine Canada's published AIAs. We report five findings: (1) Uneven compliance is observed in the completion of AIAs; (2) Reasons for automation legitimize efficiency and innovation narratives; (3) Impacts and trade-offs are framed as non-existent, positive and undermine harms; (4) Civil society organizations are non-existent in AIAs; and (5) Accountability is framed as processual mitigation of AI impacts. Despite the promise of AIAs for accountability of AI systems, our results reveal a "design-reality" gap between literature and practice. We observed that any negative impacts were framed positively; input was not elicited from the public; and an over-emphasis of self-regulation conformed to organizational procedures instead of investigating outcomes. Although submission is mandatory, its processual accountability failed to ensure compliance. We recommend strengthening accountability to include civil society, formalizing harms instead of emphasizing impacts or risks and blending processual accountability with outcomes of AI systems.

Keywords artificial intelligence · governance · public sector · policy · compliance · accountability

✉ Ana Brandusescu
ana.brandusescu@mail.mcgill.ca

1 Department of Geography, McGill University, Montréal, Canada

2 Department of Geography, School of Environment, McGill University, Montréal, Canada

Published online: 05 August 2025

1 Introduction

Algorithmic impact assessments (AIAs) have become a dominant regulatory instrument in operationalizing the effects of AI systems (Selbst, 2021). Crafted by governments and professional associations, a typical AIA comprises a series of questions on system performance, data quality, and differential impacts, producing a points-based system to evaluate the impacts and risks of AI systems. Typically voluntary, AIAs are considered important tools because they organize the way institutions interact with one another (Moss et al., 2021) and involve external stakeholders (e.g., industry, the public). A primary goal of AIAs is to establish a record of automated processes and their justifications, aiming to increase transparency and start a conversation about AI development with different stakeholder groups (Brandusescu, 2021; Mulligan & Bamberger, 2019).

AIAs have been implemented across various countries in the global north. Examples include European Commission's Assessment List for Trustworthy AI (2020), the European Law Institute's Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration (2022), Council of Europe's Human Rights, Democracy and Rule of Law Impact Assessment of AI systems (2021), the United States (U.S.)'s National Institute of Standards and Technology's AI risk management framework (2023), and the United Kingdom (U.K.)'s Institute for the Future of Work's Good Work AIA (2023). Non-governmental organizations also have developed AIAs like AI Now Institute's AIA Report: A Practical Framework for Public Agency Accountability in the U.S. (Reisman et al., 2018) and Ada Lovelace Institute's AIA User Guide for the National Health Service in the U.K. (2022). Many bodies inside and outside government have seen the value of these types of assessments, including Canada's federal government, which is our focus.

Canada's AI policies have been cited in numerous articles (Attard-Frost et al., 2024; Kuziemski & Misuraca, 2020; Metcalf et al., 2021; Selbst, 2021; Watkins et al., 2021). Canada launched the first strategy on AI in 2017, solidifying its place on the global stage. This form of governance was followed by Canada's development of a responsible AI framework spearheaded by the Treasury Board Secretariat of Canada (TBS), which included the creation of an AIA (Karlin, 2018) as part of the Directive on Automated Decision-Making (DADM), launched in April 2019. Canada's AIA is considered to be the best practice worldwide by governments and intergovernmental organizations (Darbyshire, 2022). Over six years, 26 AIAs from Canadian federal departments and agencies have been published and made publicly available. Researchers have assessed Canada's AIA instrument in detail (Gertler, 2023) and have used some AIA results to critique their underlying AI systems (Daly, 2023).

Irrespective of country, studies have emphasized the design of AIA instruments but empirical work on the usage of AIAs is sparse. We fill this design-reality gap by conducting the first in-depth review of published AIAs. We report on our comparative analysis of Canada's 26 published AIAs, which is informed by a content analysis of government documents and participant observation of AIA-focused meetings. We also include an organized repository (https://osf.io/rk8ux/) in spreadsheet form of all AIAs published between January 2020 and May 2025 to encourage further comparative research.

This paper details five findings: (1) There is uneven compliance with the AIA; (2) Reasons for automation legitimize efficiency narratives; (3) Impacts and trade-offs of AIAs are framed as non-existent, positive and undermine harms; (4) Civil society organizations are non-existent in AIAs; and (5) Accountability is framed as processual mitigation of AI impacts. More generally, AIAs promise far more than they can deliver in terms of accountability that, considering they intend to question a disruptive technology, do little to disrupt the status quo.

2 Literature review

The literature on AIAs saw a peak in 2021 with a focus on the development of the instrument and the limitations of its design. Stahl et al. (2023, p. 12,799) conducted a systematic review of the literature on AIAs and found "that the field is not yet at the point of full agreement on content, structure and implementation." According to Sloane and Moss (2023, p. 1), AIAs "are still under-defined by developers, critical scholars, and regulators, and are often conflated or referred to interchangeably." The instruments draw on varied traditions and come with numerous names: algorithmic bias assessments, algorithmic risk assessments, ethical risk assessments, and human rights impact assessments. They are evaluated with various criteria, such as levels of independence, degrees of compliance, and extent of regulation/certification (Cath, 2018; Hasan et al., 2022; Mökander & Floridi, 2021). Instruments also vary along lines of inclusion of technical (e.g., inputs and outputs) and non-technical (e.g., ethical, societal, environmental) components, internal/external organizational emphasis, inclusion/exclusion of stakeholders/public engagement, generalization and standardization, and foci on impact/bias/risk (Sandvig et al., 2014; Wernick, 2024). Hasan et al. (2022, p. 4) further distinguished audits, assurances, and assessments, arguing that "assessments are non-independent, internal, or second-party evaluations… and intended as a service to the organization." These illustrate difficulties in coming to consensus on intent and practice.

Sloane and Moss (2023, p. 2) defined AIAs as "document[ing] a range of potential effects, from privacy impacts to fairness concerns, and compar[ison of] negative impacts to the positive impacts a system might offer." Generally, the AIA literature relates impacts to harm (Ashar et al., 2024; Hasan et al., 2022; Selbst, 2021; Sloane & Moss, 2023; Watkins et al., 2021). In particular, Metcalf et al. (2021) noted that the impacts assessed in AIAs often serve as proxies for the distinct sociomaterial harms that algorithmic systems can cause. Selbst (2021) recommended that, at minimum, an AIA should standardize the types of harm that warrant compliance. Complimentary to Selbst (2021), Metcalf et al. (2021) supported the assessment of impacts to resemble harms as closely as possible.

AIA construction often is motivated by risk management, which foregrounds immediate threats to the institution's operations involving AI systems rather than broader societal impacts. Selbst (2021) studied private sector usage of AIAs but his work resonates with the public sector as well. He argued that, due to "institutional logics," companies often will prioritize liability avoidance and profitability and deploy those logics to circumvent or weaken AIA requirements (Selbst, 2021).

This can transform the industry-regulator dynamic into an adversarial one should the government's AIA rely on formalized standards and oversight, instead of a cooperative relationship that addresses harms from algorithms (ibid., p. 169). Since AIAs are often conducted internally and rely heavily on these institutional relationships, it remains uncertain whether AI-using entities will comply, especially given the limited oversight, as will be discussed.

Watkins et al. (2021) made six observations about the utility of AIAs based on their evaluation of other impact assessments, such as environmental impact assessments and privacy impact assessments. They speculated that AIAs would perpetuate the overall confusion over the term 'impact'; that reporting of impacts would likely not ameliorate harms; and mere public participation would not ensure accountability. Watkins et al. (2021) also considered what AIAs would mean for the operations and interactions of departments and agencies. Overall, they maintained a healthy skepticism of how AIAs are conceptualized and evolve during both AIA instrument development and the AI systems they are meant to assess.

Institutional relations and impacts are linked to the hope that AIAs will produce some level of accountability for AI systems. Scholars conducting research on sociotechnical systems have utilized Bovens's (2007) articulation of public accountability (Metcalf et al., 2021; Watkins et al., 2021; Wieringa, 2020). Bovens (2007, p. 9) argued that accountability is "a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences." This last clause is often eliminated from the AIA literature, one that, for Bovens (2007), makes the original formulation of accountability more about hard law (i.e., regulations) than soft law, and outcome rather than process.

However, accountability as evinced in AIAs speaks more to the processual treatment of AI systems than the outcome, with careful attention paid to procedural fairness. Procedural fairness refers to the fairness of the process by which decisions on AI design and development are made, rather than the fairness of the outcomes of those decisions. Procedural fairness emphasizes transparency, consistency, and impartiality, ensuring that all parties involved have an opportunity to be heard. In this paper, we argue that impacts exist within a particular frame of accountability that is process-based and not outcome-based. Process-based accountability, particularly regarding AI systems, is internal. It prioritizes the procedures to achieve a specified system objective. Whereas process-based accountability ensures that processes comply with an organization's ethical and legal requirements; outcome-based accountability ensures that these actions achieve the desired results (Lee et al., 2019; Patil et al., 2014). At minimum, authors have noted that a feedback loop should be established between process and outcome (Patil et al., 2014). Finally, some researchers have implied that the 'road to hell is paved with good intentions', where "process accountability can thus spiral downward into blind conformity that sustains deficient decision practices" (Patil et al., 2014, p. 11). This form of internally-focused process accountability often lacks outside scrutiny that is essential for a robust framework of accountability (Metcalf et al., 2021).

Transparency is usually considered a precondition for accountability, whether that accountability is process or outcome-based (Diakopoulos, 2020). Despite the many promises that transparency of decision-making processes furthers inclusion, legitimacy and accountability as well as reduces harm in AI systems (e.g., Waldman & Martin, 2022), authors warned about a hyper-focus on transparency. Ananny and Crawford (2018) suggested that transparency places an excessive burden on the individual to understand and be able to act on that material; transparency assumes that the accumulation of "facts" leads to change. Moss et al. (2021, p. 9) argued that "voluntary commitments to… transparency do not constitute accountability" and represented a soft form of accountability (Fox, 2007). As an antidote to these challenges, Diakopoulos (2020) proposed a "usable transparency" that moved beyond abstract principles and values towards human-centered approaches that align with stakeholders' abilities to understand and engage with the information.

More specific to the instrument, Waldman (2019, p. 624) saw fundamental flaws in the processual accountability mechanisms of AIAs, meaning that proponents supported a "system that not only lays the groundwork for algorithmic decision-making but sees its proliferation, despite its biases, errors, and harms, as a good thing." For Waldman (2019), AIAs as designed can cause more harm than resolve them.

To address some aforementioned shortcomings, Stahl et al. (2023) concluded with two important elements for an AIA: AIAs should capture whether the AI system is expected to have a variety of social impacts and which stakeholders are likely to be impacted. Because delineating impacts represent institutional priorities and politics, impacts become "site[s] of contestation shaped by social, economic, and political power" (Watkins et al., 2021, p. 1014). Indeed, "a preoccupation with procedure may render other important aspects of involving the public invisible" (Watkins et al., 2021, p. 1017). This aspect of public involvement can differentiate Canada from other countries using AIAs. Compared to the U.S., there is no statutory, constitutional or charter right to public participation in Canada. Instead, government relies on interest groups to represent the public even as they "tend to be comprised of a socio-economic elite" (Smith, 1984, p. 255). This can limit civil society's ability to marshal significant political power to counter the outsized influence of industrial AI systems development (Sieber, 2022).

Many reviews of AIAs cite Canada's AIA as exemplar, although authors focus on instrument design. For example, Selbst (2021) labelled Canada's AIA an "efficient" and "thoughtful" but a "quite general" questionnaire with numerous drop downs and yes/nos. Gertler (2023, 2024) examined the process of completing the questionnaire, including the creation of an accompanying guide.[1] He concluded that the "AIA tool is largely performative, yet its significance marks an important tactical intervention" (Gertler, 2024, p. 41). In mapping AI use in the Canadian government, Daly (2023, p. 88) categorized public sector AI systems by agency-specific purpose, among them "enhancing the accessibility of public-facing resources", "using information to create and enhance models of natural and human activity", and "managing enforcement resources." He proceeded to link these reasons for automation to potential utopian or dystopian visions of government (Daly, 2023), which is somewhat beyond our practical orientation.

[1] https://aia.guide/

Unlike the literature to date that continues to centre instrument design (or, more recently, on the practitioners who fill out the AIAs (Ashar et al., 2024)), we focus on the outcomes of the instrument, that is the AIA results. An examination of the answers of Canada's AIAs could close the instrument's design-reality gap and help reduce the uncertainty over the AIAs' revealed purpose.

3 Methodology

Canada's AIA is a points-based questionnaire to assess impact and risk mitigation of the development and deployment of automated decision-making systems (hereafter "AI systems"). The AIA was launched by TBS in March of 2019. The earliest published AIA was in January 2020 and the latest AIA was published in May 2025 (as of June 2, 2025). The AIA questionnaire has gone through ten versions; only Versions 0.9.1 and 0.10.0 have been used (hereafter referred to as V9 and V10).

Canada's AIA scores the impact of an AI system. Figure 1 shows the classification of the 85 questions in V10 (V9 contains 60 questions). V10 is divided into 51 risk questions (maximum score 126) and 34 mitigation questions (maximum score 46). Some boxes in Fig. 1 refer to multiple questions. The higher the total score, the higher the risk of the AI system. The AIA contains four impact levels, with the final score based on comparisons between the two scores. If the mitigation score is less than 80% of the possible total then the raw impact score is used. If the mitigation score exceeds 80% then the raw impact score is reduced by fifteen percent (Government of Canada, 2024a). Presumably "effective" mitigation reduces perceived risk. More importantly, an impact level determines the amount of actions required of the department or agency. Level 1 requires "explanation"; "peer review" is required above Level 1 (e.g., from "Qualified expert from a federal, provincial, territorial or

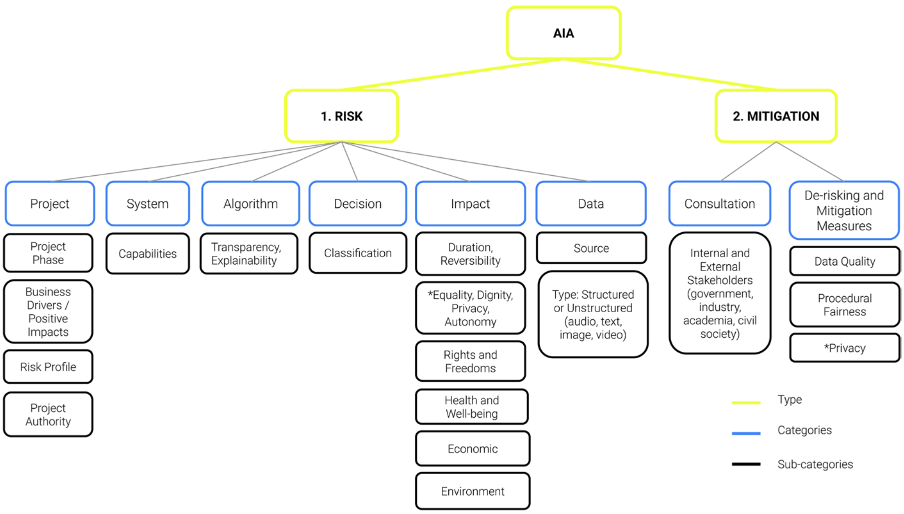

Fig. 1 Visualizing the structure of Canada's AIAs 85 questions. Asterisks indicate additions to subcategories only appearing in V10

municipal government institution”). At higher levels, additional reports must be produced (Levels 2–4); Level 4 requires permission from TBS to operate the AI system. The TBS director responsible for the AIA, “offered up as a Level 1 decision whether someone is eligible for a $3 rebate on an energy-efficient lightbulb. At Impact Levels 3 and 4, it could be things like getting a visa for coming to Canada or imprisonment decisions, things of that nature” (Reeveley, 2021).

Scoring is considered a mechanism to ameliorate negative impacts. Kaye (2024) cautioned that the urge of AIA developers to quantify impacts could potentially increase risk because the numbers would strip impacts of context and nuance. We took this as an opportunity to supply that nuance, by analyzing each completed AIA in detail.

We employed three methods. First we conducted a content analysis of federal government policy documents related to the AIAs, including the Directive of Automated Decision-Making (DADM) (Government of Canada, 2023), which is the overarching Directive that guides the development and evolution of the AIA. To guide this analysis, we borrowed from Hsieh and Shannon’s (2005) inductive coding and iterative category development that also have been applied to research on AI governance in Canada (Attard-Frost et al., 2024). Content analysis helped generate our interpretation of the description of the AIA results (see Appendices 2 and 3).

Our second method was participant observation vis-à-vis the hosting of two public meetings (https://osf.io/rk8ux/) with experts from civil society and government involved in the DADM and AIA. Musante and DeWalt (2010, p. 92) view participant observation as a method that can “develop a holistic understanding of the phenomena under study that is as objective and accurate as possible.” We found this method improved understanding of the actors and their relationships with stakeholders in the AI policy ecosystem in Canada.

Lastly, we conducted a comparative analysis of Canada’s AIAs using Luna-Reyes et al.’s (Luna-Reyes et al., 2010) method on digital government policy in North America. Comparative analysis seeks to reveal changes, similarities, or inconsistencies in policy development, which helps in assessing the alignment of policies across different departments and agencies as well as timeframes.

Twenty-six AIAs were uploaded to Canada’s Open Government Portal from January 2020 to May 2025 (Government of Canada, 2025).[2] The AIAs were published using two versions of the instrument, V9 (Government of Canada, 2022) and V10 (Government of Canada, 2024b). To conduct our comparative analysis, we extracted the answers for the English versions of the AIA (a combined 316 pages of text). The AIAs range between eight and 24 pages, with a median of nine for V9 and twelve for V10. There are 33% more pages now with V10, although V10 also contains 42% more questions than V9 (as shown in Fig. 1). We entered the responses in two open datasets, which we made publicly available at https://osf.io/rk8ux/.

[2] Note that when one searches “algorithmic impact assessment” in the open government portal, it will generate 27 results, because the AIA itself also is included in this list. Since it is not an AIA, we did not count it in our review.

4 Findings

Published AIAs uncovered considerable variation in AI systems, ranging from Veteran Affairs Canada (VAC)'s Automation Development to Support Disability Benefit Decision Making and Transport Canada's Pre-load Air Cargo Targeting (PACT) Program to Immigration, Refugees and Citizenship Canada (IRCC)'s Automation Tools to Help Process Privately Sponsored Refugee Applications; from the Royal Canadian Mounted Police (RCMP)'s Griffeye Tool for the detection of child sexual abuse material to Canada Border Services Agency (CBSA)'s Client Reporting and Engagement System/ReportIn facial recognition technology for the collection of facial biometrics and accurate location. This range of responses allowed us to derive five findings that are likely generalizable for future AIAs in Canada and in other countries.

4.1 Uneven compliance is observed in the completion of AIAs

Canada's AIAs are mandatory for federal departments and agencies implementing AI systems (Institute for the Future of Work, 2022). The TBS Guide on the DADM explains that "AI systems involved in administrative decision-making are required to comply" with an AIA (Government of Canada, 2024c). That encompasses "any decision [or partial decision] that has the potential to affect legal rights, privileges or interests" (ibid.). The category of administrative decision-making is expansive, including vital decisions that allow or deny people entry into Canada or affect hiring processes. The Starling Centre (2024) launched a public registry of automated decision-making systems in the Canadian federal government, finding a total of 303 AI systems (as of February 1, 2024).[3] Presumably we should have found hundreds of AIAs. However, Canada published only 26 AIAs in six years. Departments and agencies may assume compliance is unnecessary because their AI systems are uninvolved in this type of decision-making but this likely does not describe 91% of institutional AI systems without an AIA.

In many countries, departments and agencies related to the security state are exempt from compliance with AI-related regulations and policies. In Canada, the Canadian Security Intelligence Service is exempt. These exemptions can transcend national security. According to the first public meeting on the DADM and AIA (https://osf.io/rk8ux) organized by AI researchers and policy experts, participants were surprised that the Canada Revenue Agency (CRA), the largest government employer in Canada and one that deploys multiple AI systems, also is not covered by this Directive.[4] Instead, CRA developed its own version of the AIA, which is not publicly available (e.g., not published on Canada's Open Government Portal). Conversely, National Defence, Service Canada, Environment Canada, and Natural Resources Canada are covered by the DADM, but have no published AIAs. Interestingly, the national police force, the RCMP, has (see Fig. 2). CBSA only recently published its

[3] https://tagcanada.shinyapps.io/register/

[4] A 1999 CRA Act (Section 30) excludes the CRA from TBS administrative policies: https://laws-lois.justice.gc.ca/eng/acts/c-10.11/page-2.html#h-49799

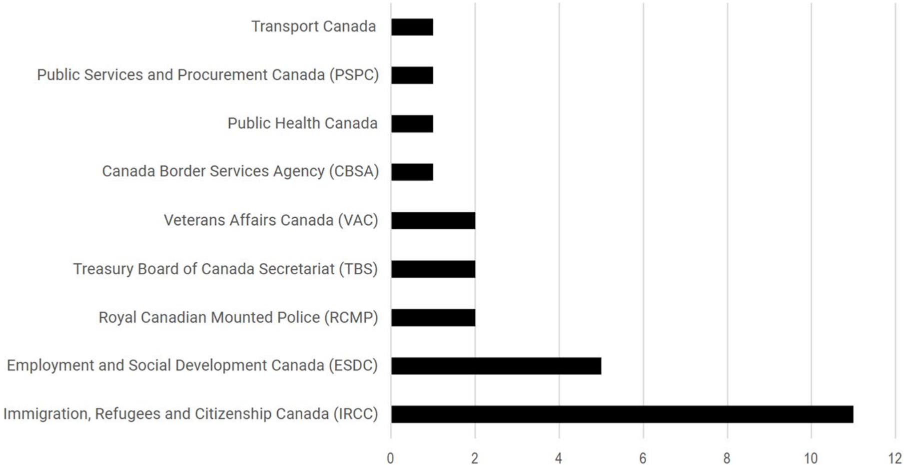

Fig. 2 Total number of published AIAs by the Canadian federal government departments and agencies as of May 2025. The detailed list is in Appendix 1

first AIA but claimed its system as a trade secret and therefore not subject to significant disclosure (Christian, 2024).

As suggested above and in Fig. 2, only a fraction of the departments and agencies have complied with the AIA. Indeed, almost half of the AIAs have been published by IRCC. We speculate that IRCC has submitted the most AIAs because of the AIA question that asks whether the AI system supports “an area of intense public scrutiny (e.g., because of privacy concerns) and/or frequent litigation”; IRCC answered “Yes” about intense public scrutiny/litigation for eight of eleven of its AIAs.

Certain high-risk AI systems being used by departments and agencies are known for harmful impacts. CBSA uses facial recognition at airport kiosks to verify traveller identities (Karadeglija, 2024). Karadeglija cited Redden of the Starling Centre, who highlighted numerous concerns, such as wrongful arrests linked to facial recognition in the U.S. Similarly, the Canadian Immigration Lawyers Association warned that CBSA’s plan to use facial recognition for another app tracking deportees poses serious surveillance risks (Christian, 2024). The Associated Press reported that the U.S. has weaponized a similar report-in tool to streamline deportations (Gonzalez & Salomon, 2025). Given the sharing of refugee and asylum-seeker data between Canada and the U.S., the use of the app heightens the potential negative impact.

4.2 Reasons for automation legitimize efficiency and innovation narratives

The AIA supplies a pulldown menu to report reasons to automate operations. These include “Lower transaction costs of an existing program”, “Existing backlog of work or cases”, “Improve overall quality of decisions”, and “The system is performing tasks that humans could not accomplish in a reasonable period of time.” Both V9 and V10 include in their pulldown menu an open ended explanation titled “Other (please specify)” to which 12 of the 26 departments/agencies responded. Our content analysis helped unpack and expand on the reasons listed in the AIAs. We created our own categorizations to augment the pre-prepared list (see Appendix 2 for further details). Most reasons were related to efficiency, such as "Find, optimize efficiency in-house and for customers", "Reduce redundancy, bureaucracy", "Reduce backlog", and "Save on costs." Innovations such as "Shift employees to higher level decision-making roles" and "Respond to big data" also were implied. The term "efficiency," which is not explicitly mentioned in the pulldown menu, remains intertwined with innovation, especially within the public sector (Mikhaylov et al., 2018). Here, innovation often seems more compatible with achieving efficiency rather than generating new insights. The "Encourage economies of scale" rationale encapsulates a dominant discourse of AI. In pursuit of increased efficiency, delivering services at scale promises greater productivity, lower costs, rapid growth, and centralized control, all of which is attractive to agencies seeking to exploit a transformative technology.

Reasons beyond efficiency included "Reduce complexity", "Allow experimentation", "Limit scope", and "Trust vendors." "Increase objectivity" aligns with "Increase integrity, legitimacy, and trust of government programs" with the former emphasizing the government's trusting AI systems and the latter focusing on fostering public trust in government itself. Some outliers like "Respond to time-sensitive issues," supported crisis management and "Enable fairness and diversity" reflected equity. Other categories spoke to harm, such as "Protect the children" and "Reduce toxicity of content moderation."

"Improve accuracy" as a category combines efficiency and objectivity, for example by reducing employee error in transcription (RCMP's Voice 2 Text application), which suggests the benefits of removing humans from the loop at certain stages. Even this prosaic task, however, could inject bias should the AI system fail to recognize all accents and speech patterns. "Augment identification" via biometrics may appear similarly benign, but new forms of personal identification have been found to amplify bias, demanding at minimum some estimation of that bias (Drozdowski et al., 2020).

We continue to note an absence of harm reduction as a reason to justify the AI system. This is converse to the general sentiment of meetings on the DADM and AIA in 2021 and 2022, where participants focused their attention on harms. We see, however, an extensive focus on risk. For example, Transport Canada's AIA "aims to identify and apply risk mitigation measures to inbound high-risk air shipments that could contain concealed improvised explosive devices or other threat items prior to loading and departure to Canada." Only IRCC's Spouse and Common Law Partners AIA and RCMP's Griffeye Tool AIA noted "Yes" to impact on (i.e., risk to) employees. Griffeye offered "a reduction in the repetitive traumatic exposure by investigators to such media will benefit their well being." Neither transparency nor accountability appear as reasons in the responses.

Many reasons chosen for automation are determined by the design of the AIA. Submitters are asked to retroactively justify their AI systems with reasons from the pulldown menu. These choices beg potential answers, suggesting a degree of social desirability bias on the part of the instrument designer. It sets the stage to frame AI adoption as invariably positive.

4.3 Impacts and trade-offs are framed as non-existent, positive, and undermine harms

The essence of AIAs is to assess impacts. Canada's AIAs cover a range of impacts resulting from AI systems, which include rights and freedoms, equality, health and wellbeing, economic, and environment. AIA V10 adds a sixth impact, to individual equality, dignity, privacy, and autonomy. Ninety-seven percent of submitters reported that their AI system produced (in the implementation stage) and anticipated (in the design stage) little to no impact across all five questions.[5] Only two answers were "moderate impact" (i.e., GC HR and Pay - Leveraging AI for backlog reduction, Integrity trends analysis tool); three answers labelled the AI system as having a "high impact" (i.e., Mental Health Benefit app, Automation Development to Support Disability Benefit Decision Making {the latter replied to high impact in 2 different categories}). All instances of moderate impact or high impact, except one (i.e., Integrity trends analysis tool), were related to health and well-being. For example, most AIAs answered that "The system is expected to have little to no negative impact on the health and well-being of individuals as it will be used to triage applications and to automate certain positive eligibility determinations."[6]

Despite more than half of the AIAs reporting their AI systems would be subject to intense public scrutiny, "low impact" was listed in the majority of the answers to the impact questions. AI systems were considered low impact, even though almost one third (8 out of 26) were labelled as part of high stakes decision-making situations. The two VAC systems were listed as "high stakes" but only were reported as high impact in one of the five questions for the Mental Health Benefit AI system (V9). Two of six questions (V10) were reported as "high impact" for the Automation Development to Support Disability Benefit Decision Making AI system. Another concerning revelation is that almost one third of the AIAs (8 of 26) were listed as having "particularly vulnerable" clients. With the exception of VAC's ("high impact") and PSPC's ("moderate impact") submissions, the systems were listed as having "little to no impact."

One notable case was VAC's Mental Benefit AI system. VAC did not answer the equality, dignity, privacy, and autonomy question for this particular system. Skipped questions are difficult to detect unless one conducts a thorough comparison of each AI system to the question answered in each AIA response. Submitters could skip questions and the reader would never know from the generated PDF. TBS alerted us to this ongoing issue (AI and Data Director, personal communication, August 28, 2024). According to the Canadian government's GitHub on the AIA:

[5] We counted the responses from the two AIA versions: 139 of 144 answers in the 26 AIAs (1 question was left blank for 1 AIA submission).

[6] This particular example pertains to IRCC's AI systems, which often supported triaging immigration and visa applications to the department.

> There is no validation in the AIA. It is possible to skip questions or entire sections and this is not reflected in the results page. The answers on that page only show the questions that were answered.[7]

In addition to reporting "little to no impacts," departments and agencies appeared reluctant to acknowledge any negative impacts of their systems. This could be due to the risk-averse culture historically within the Canadian federal government (Public Policy Forum, 1998). Risk-adversity can be extended to fear of punishment should AI systems be reported as causing harm or being viewed as a negative to the organization. Invariably, submitters 'spun' impacts as positive. Consider PSPC's AIA, which reported that "it is anticipated that the decision's impact on the health and well-being of individuals will fall within the moderate range, bringing improvement to both the work environment and the quality of worklife for CAs [compensation agents]." Later, PSPC acknowledged potential unnamed trade-offs made in the agency's AI system while extolling the benefits "that serve… both dedicated CAs and public servants impacted with pay issues." A TBS trade-off example highlights a similar conclusion: "By prioritizing cases [via AI] that have a significant impact on claimants or the EI program, officers can allocate their attention more effectively to critical tasks, resulting in improved service delivery." The VAC responded that "Mental health benefits impact client wellness by providing early access to coverage for required treatments." CBSA's AIA reported "little to no impact" on the environment while adding a positive impact in the details "due to less travel and transportation effects". The overwhelming "little to no impact" answers of this and other AIAs validates our analysis that the government views impact as uniformly negative.

We also found significant copy-pasting. Many examples from IRCC have nearly identical answers, reported across numerous impact questions for one AI system and across many AI systems. This suggests the burden of compliance and possible reluctance to reflect on the impacts, positive or negative.

Unlike the AIA submitters, researchers and practitioners tended to conceptualize impacts as uniformly negative (e.g., Watkins et al., 2021) because they focused on impacts of greatest concern. Schneider (2024) interviewed researchers with regards to VAC's Mental Health Benefit AIA, who cautioned against triaging files and automating determinations; the journalist reported that the agency was insufficiently diligent in explaining the decisions generated by VAC's AI system. Nonetheless, AIA submitters expressed that stakes could be high *and* the impacts could be positive. We perceive this as a prime challenge in connecting impacts to harms if departments and agencies seek to blunt criticism of AI by foregrounding the positive aspects of developing a system at the expense of the harms.

In a search for harms in accordance with the literature, we combined our interpretations in Appendix 3 with responses from the trade-off question of the AIA, a question introduced in V10. We found that efficiencies could eclipse fairness; human bias could be replaced by machine bias; human rights could be reduced to a UI issue; and training and quality control could ameliorate negative impacts. Concerned about impacts, the text of IRCC's Administrative Activities Support Tool AIA contained

[7] https://github.com/canada-ca/aia-eia-js/issues/1124

constant reassurance that the AI system would neither replace human decision-making nor eliminate employment. Potential violations of individual privacy due to AI systems were invoked in IRCC's Family Class Spouses and Partners Overseas AIA. One odd outlier is a concern that at-risk individuals would be overrepresented in ESDCs Old Age Security Extended Absence Leave Prioritization AI system. In certain cases, no impact would be experienced because the private sector could take its business elsewhere (Transport Canada) or that compliance was only voluntary (CBSA). Similar to impacts, trade-offs also were characterized as positive. Concerns were ameliorated by the assertion that a government employee would always be in control or a human would always be there to take care of a client. Interestingly, retaining a human-in-the-loop can be framed as both negative–an introduction of bias, which would be reduced by the AI system–and positive–"do not worry, the humanity of government will be preserved."

The positivity discourse lies in stark contrast to expert reflections. In the second public meeting on the DADM and AIA (https://osf.io/rk8ux), a participant noted that

> From a social equity perspective, the ArriveCAN [Covid-19] app does not do justice to travellers coming from some countries where proofs of vaccination are not always "scannable." Not every traveller will show up with a PDF file that you can enter into the app on the ArriveCAN app. This seems to perpetuate the inequities that are already a common occurrence with many AI systems. Systems tend to discriminate against some countries as well as some groups of people. From a social equity perspective, this is a concern.

It is notable that the question on trade-offs does not increase the score, which means this potential negative impact does not implicate the need for mitigation.

4.4 Civil society organizations are non-existent in AIAs

Participation of civil society and impacted publics with public sector AI is considered crucial for improved public service delivery and democracy vis-à-vis AI development and deployment (Reisman et al., 2018; Sieber et al., 2025). A set of questions called "consultations" asks submitters to describe the internal and external stakeholders engaged in the development of the AI systems. A separate question asks about the type of external stakeholder; the list for types of internal stakeholder is longer (e.g., "Strategic Policy and Planning", "Legal Services", "Access to Information and Privacy Office"). Table 1 shows the total amount of external stakeholders consulted for each group.

Irrespective of civil society representation, Table 1 highlights the unevenness of reporting. Over half of AIA submitters (15 out of 26) replied "Yes" to consultations but left the description of external stakeholders empty (we labelled these cells as "None mentioned" in Table 1). Six entries listed no consultations of any external stakeholder groups, which suggests a disincentive to engage externally. If an AIA submitter reports engagement with stakeholders then they are assigned 1 point; if an AIA submitter reports engaging both internal and external stakeholders then they are assigned 2 points, which may require mitigation. Submitters thus encounter a

Table 1 AIA answers from the external consultation questions. External stakeholders consulted in AIAs are organized in alphabetical order by government department or agency. Answers reveal that external stakeholder groups were often not mentioned. Question marks denote that a stakeholder group was consulted but no details were provided on specific individuals or organizations. Professional associations are categorized as industry

Department/ Agency	AI System	External Stakeholder Groups			
		Government	Industry	Academia	Civil Society
CBSA	Client Reporting and Engagement System (CRES)/ReportIn	?	None mentioned	None mentioned	?
ESDC	Employment Insurance Machine Learning Workload	None mentioned	None mentioned	None mentioned	None mentioned
ESDC	Old Age Security Extended Absence Leave Prioritization	None mentioned	None mentioned	None mentioned	None mentioned
ESDC	Record of Employment Comments Assessment Using AI	None mentioned	None mentioned	None mentioned	None mentioned
ESDC	Reducing Employment Insurance Backlog	None mentioned	None mentioned	None mentioned	None mentioned
ESDC	Repayment Assistance Plan–Enhanced Verification Model	?	?	None mentioned	None mentioned
Health Canada	ArriveCAN Proof of Vaccination Recognition	5	76	None mentioned	None mentioned
IRCC	Administrative Activities Support Tool	None mentioned	None mentioned	None mentioned	None mentioned
IRCC	Advanced Analytics Triage of Overseas Temporary Resident Visa Applications	2	?	?	None mentioned
IRCC	Advanced Analytics Triage of Visitor Record Applications	1	None mentioned	?	None mentioned
IRCC	Automate the review of non-complex applications for Temporary Resident Visas and Work Permits made under the Canada-Ukraine Authorization for Emergency Travel	None mentioned	None mentioned	None mentioned	None mentioned
IRCC	Automated Triage and Positive Eligibility Determinations of In-Canada Work Permit Applications	?	?	?	None mentioned
IRCC	Automation Tools to Help Process Privately Sponsored Refugee Applications	1	None mentioned	None mentioned	?
IRCC	Family Class Spouses and Partners Overseas Applications	None mentioned	?	None mentioned	None mentioned
IRCC	Integrity Trends Analysis Tool	1	None mentioned	None mentioned	None mentioned
IRCC	International Experience Canada Work Permit Eligibility Model	None mentioned	None mentioned	None mentioned	None mentioned
IRCC	Passport Automated Decision-Making in IRCCs Global Case Management System (GCMS) - Passport Program Modernization Initiative	1	None mentioned	None mentioned	None mentioned

Table 1 (continued)

Department/ Agency	AI System	External Stakeholder Groups			
		Government	Industry	Academia	Civil Society
IRCC	Spouse or Common-Law Partner in Canada Advanced Analytics Pilot	1	None mentioned	None mentioned	None mentioned
PSPC	GC HR and Pay - Leveraging AI for backlog reduction: AI assistant prototype	None mentioned	?	None mentioned	None mentioned
RCMP	Griffeye Tool	?	?	?	None mentioned
RCMP	Voice 2 Text application	None mentioned	None mentioned	None mentioned	None mentioned
TBS	ATIP Online Request Service	2	?		?
TBS	Using AI to automate candidate evaluations in the staffing process's assessment phase	None mentioned	None mentioned	?	None mentioned
Transport Canada	Pre-load Air Cargo Targeting Program	None mentioned	60	None mentioned	None mentioned
VAC	Automation Development to Support Disability Benefit Decision Making	None mentioned	1	None mentioned	None mentioned
VAC	Mental Health Benefit	1	?	None mentioned	?

perverse incentive to not engage with nor report external stakeholders. There are two exceptions to this practice. Health Canada and Transport Canada were explicit in naming entities and listed a significant variety of external stakeholders, albeit overwhelmingly from industry. Conversely, we found zero examples of stakeholders consulted from civil society as well as academia, the latter of which some agencies and departments viewed as proxies for civil society.

One should examine external stakeholders carefully. In Canada's AIAs, external stakeholders can include government departments and agencies (an odd category placement for non-government readers like ourselves), such as Statistics Canada and the Office of the Privacy Commissioner. A related observation is that the DADM refers to non-governmental organizations (NGOs) (Government of Canada, 2023), which could induce ambiguity around the NGOs' exact inputs into the AIA. NGOs could encompass professional associations, business associations, unions or "Qualified researchers from a relevant non-governmental organization." A large number of professional associations appear in Table 1,[8] which were classified as industry; their prevalence suggests a consultation that is expert-driven and potentially alienating to civil society (Attard-Frost, 2023; Brandusescu & Sieber, 2025). Because AIA submissions are self-reports provided by submitters, a department or agency could determine how stakeholders are classified and who has a stake in the AI system. However, not all stakeholders are equal; they possess differential access to resources and power

[8] The list of names are available are https://osf.io/rk8ux

bases (Sambuli, 2021). Should representatives of civil society be involved, then departments and agencies should account for unequal dynamics among stakeholders.

In addition to the lack of detail in the external stakeholder lists, no mitigation actions require civil society or public engagement. Low to very high impact AI systems (Impact Levels 2–4) *may* require consultation from NGOs but also could be satisfied by qualified experts from government or the aforementioned "Qualified researchers from a relevant non-governmental organization." As above, these experts do not translate to representatives of civil society.[9]

Regarding the openness of the AIA process and the opportunity for public engagement, a government official confirmed that

> There are [only] minimal requirements around this. The only engagement which is required is the peer review requirement, which can include members from universities, for example, professors acting as consultants. Sometimes TBS is engaged, sometimes they are not.

Presumably medium to high impact AI systems would include AIAs with civil society representation (e.g., impacted and vulnerable communities) in their peer review. CBSA reported consultation with international organizations and civil society; however, they listed no specific organizations for one to assess the degree of accountability. Similarly, the VAC did not disclose impacted client groups in its Mental Health Benefit AI system, which could verify its reported improvements to "client wellness." RCMP's AIA noted that only positive impacts would result from their AI system, even in high impact cases. Its AIA also did not include civil society (e.g., children's aid organizations) in its stakeholder engagement. Unfortunately, the AIA could enable 'gaming the system', since the AIA instrument requires no civil society engagement at any level of impact or mitigation and many submitters did not appear obliged to identify specific organizations.

4.5 Accountability is framed as processual mitigation of AI impacts

Sections 4.3 and 4.4 addressed responses related to impacts. Submission of an AIA is only recommended twice during the design and development phases of the AI system. These two snapshots assume unforeseen negative impacts like biases can be identified early and mitigated.[10] IRCC suggests the premature nature of completing the AIA in their Automated Triage and Positive Eligibility Determinations of In-Canada Work Permit Applications as well as in their International Experience Canada

[9] https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appC

[10] "3.1 When to complete the AIA: The AIA should be completed at the beginning of the design phase of a project. The results of the AIA will guide the mitigation and consultation requirements to be met during the implementation of the automated decision system as per the directive; The AIA should be completed a second time, prior to the production of the system, to validate that the results accurately reflect the system that was built. The revised AIA should be released on the Open Government Portal as the final results; Reviewing and updating the AIA: The AIA should be reviewed and updated on a scheduled basis, and when the functionality or scope of the system changes (subsection 6.1.3). The schedule of review can be aligned with and informed by the monitoring (subsection 6.3.2) and reporting requirements (subsection 6.5.1)".

Work Permit Eligibility Model AI systems. When one evaluates an AI system is key. Following Moss et al. (2021) who argued that the questionnaire-based nature of AIAs failed to connect the outcomes of processual systems with actual harms, a processual accountability could explain the challenge in finding negative impacts. In other words, impacts were anticipatory but not actual. In this section, we examine the consequences of using a processual model for accountability.

Procedural fairness is the epitome of processual accountability and a focus of Canada's AIA. With 17 questions, procedural fairness is the largest subsection in "de-risking and mitigation measures" of the AIA.[11] The DADM defines procedural fairness as "A guiding principle of governmental and quasi-judicial decision-making [that] increases or decreases with the significance of that decision and its impact on rights and interests." Overall, department and agency answers for procedural fairness questions were resoundingly positive. The submissions reassured readers that their processes had accounted for or would account for fairness. For instance, IRCC asserted that "All of the tools' rules are carefully vetted for potential bias. The tools will also undergo a GBA [Gender Based Analysis] Plus assessment."

Important to procedural fairness is the concept of proportionality. The more significant the decision's impact on the claimant, the higher the obligation for procedural fairness required. Departments and agencies must implement procedures to assess, reduce, and manage algorithmic bias and its disproportionate effects (Heisler, 2022, p. 37). Although procedural fairness plays a major role in the AIA instrument, should a department or agency mark an impact as low to none (see Sect. 4.3) then the need to provide proportional mitigation is reduced because the score is reduced.

Three AIA questions loosely correspond to oversight within the procedural fairness section of the instrument: "Will there be a recourse process established for clients that wish to challenge the decision?"; "Will the system enable human override of system decisions?"; and "Will there be a process in place to log the instances when overrides were performed?." For the question on recourse, several departments and agencies, such as RCMP (Griffeye Tool, Voice 2 Text application), VAC (Mental Health Benefit application), and PSPC (AI for backlog reduction), answered "No." These same agencies responded "Yes" to the two human override questions, suggesting an implicit reliance within government on a 'trust us' approach, where human oversight is sufficient when internally rather than when externally accountable. This suggests a reluctance to engage with external mechanisms for oversight, preferring internal control over the review processes.

Transparency functions as a key precondition to processual accountability of AI systems, represented in the AIA as the availability of various accompanying documents and procedures. More specifically, four questions pertain to whether documentation is publicly available. Every department and agency answered "No" to all questions with one exception, PSPC. We checked this exception from PSPC and found no publicly available documents. Rather than transparency, what we found was opacity, even though Canada's AIA assumes "Processes to ensure data is representa-

[11] "Data quality" is the second largest subsection in "de-risking and mitigation measures" of the AIA, with 10 questions and "privacy" is third, with 7 questions.

tive and unbiased, as well as transparency measures related to those processes."[12] Interestingly enough, we discovered references to transparency in the impacts section but not in the mitigation section. For example, IRCC's International Experience Canada Work Permit Eligibility Model's impacts on the equality, dignity, privacy, and autonomy of individuals reported that "To ensure fairness and transparency, the system's rules are based only on data elements with a clear link to legislative, regulatory and contractual requirements." This lack of commitment to making documents publicly available is somewhat ironic for Canada that has been committed for over a decade to the open-by-default principle[13] and which hosted the Open Government Partnership Global Summit in 2019.[14]

An exception to the purely processual was seen in CBSA's ReportIn AIA attempt to identify and remediate any bias (see Appendix 3). CBSA utilized Amazon, who hired a third party firm to conduct an assessment of their system, Amazon Rekognition, that is being used for the ReportIn AI system.

More broadly, no external body provides independent oversight or tracks the consistency of processual protocols referenced in documents. The agency in charge of collecting and storing the AIAs, TBS, does not track the mitigation (AI and Data Director, personal communication, August 28, 2024) nor the need to submit AIAs. Oversight is completely internal and voluntary.

5 Discussion

Canada's AIA is heralded as the gold standard of AI assessments. Its AIA is mandated by the six-year old DADM; however, only 26 AIAs have been published throughout the federal government. Although the policy is mandatory, TBS has delegated authority to departments and agencies to determine which AI systems warrant disclosure via reporting AIAs and whether AI-developing entities (e.g., companies, research institutes) need to comply. The findings suggest that this approach to self-regulation does not align with researchers' expectations, particularly regarding meaningful public involvement and transparency.

Our research provides empirical support for the observations found in Watkins et al. (2021) about impacts and harms. Converse to the contention that impacts would elicit negative effects, impacts were invariably characterized as positive and reported as producing "little to no effect." If the majority of AI systems can be characterized as "little effect" then this suggests little need for accountability, processual or otherwise. This minimization of harm and the positive spin on impacts were the most surprising to us. The published AIAs also could represent selection bias if more harmful AI systems were excluded from AIA reporting. Alternately, AI systems were possibly introduced into an already contentious process so the systems were perceived to neither add nor decrement existing potential harm.

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

[13] https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=28108

[14] https://www.opengovpartnership.org/events/ogp-global-summit-2019-ottawa-canada/.

Certain contradictions in responses could likewise lie in the conflation of impacts with risk, stakes, and public scrutiny. The implication of the conflation is that almost all impacts could be reduced to institutional risk (Hasan et al., 2022). Moreover, departments and agencies could decide what constituted an impact or risk in an administrative decision. They also could decide whether the system was making a decision. For example, IRCC reported that the triaging system was not actually rendering a decision; therefore, it did not induce any potential harm. Unfortunately, there is little to no external oversight on algorithmic decision-making, which for AIAs, should represent an integral check on administrative decision-making.

Researchers of AIAs and AI ethics recommend participation of stakeholders, which should include vulnerable and marginalized communities who are impacted by the AI systems (e.g., Sieber et al., 2024; Stahl et al., 2023). In practice, neither these types of publics nor their proxies were reported in the AIAs. For instance, veterans appeared not to have been consulted in the VAC case (Schneider, 2024). Gertler (2024) asserted that Canada does not necessarily prioritize engagement of the public or potentially impacted communities in similar ways as other AIAs are designed to do (Reisman et al., 2018). Unfortunately, the lack of public engagement in AIAs follows a pattern consistent with Canada's recently proposed and failed AI regulation (Brandusescu & Sieber, 2025).

Of the AI systems for which AIAs were submitted, ArriveCAN received the most public scrutiny in the press and in government investigations.[15] ArriveCAN was a joint project of Health Canada and CBSA but only Health Canada submitted the AIA. Had CBSA submitted an AIA related to ArriveCan, it might have realized the value of public consultations with elderly or people without mobile phones crossing the US-Canada border. Border crossings typically suppress mobile phone, Wi-Fi, and Internet connections for security reasons but that meant the elderly or people without mobile phones crossing the US-Canada border were directed by border agents to places like libraries to use the app as well as support to comprehend the app (Quenneville, 2022).

We observed an over-emphasis on efficiency and innovation as drivers of AI adoption, which reflects a long-standing shift in the public sector towards a market-orientation. This neoliberal perspective also helps explain the potential failure of the processual. Waldman (2019, pp. 628–9) argued that the same neoliberalism, which gave rise to AI-augmented decision-making, has reduced accountability to mere compliance. Even when accountability remains processual, this shift can weaken processual guardrails that regulate AI systems, as corporate interests may co-opt processes and obscure any erosion of principles like fairness, equality, and human dignity, and dissuade users and policymakers from pursuing stronger reforms once the processual requirements are met. Published AIAs demonstrated a further market alignment within government.

Processual accountability was evident throughout the AIAs. Responses to procedural fairness, a key pillar of processual accountability, were overwhelmingly positive but required no independent oversight to assess the AI system's processes. Responses also did not require outcome-based accountability for the AI system's

[15] https://www.priv.gc.ca/en/opc-news/news-and-announcements/2024/an_240319/.

impacts. Researchers of AIAs like Stahl et al. (2023), Watkins et al. (2021), and researchers of accountability like Bovens (2007) stressed the importance of relationships as the foundation of accountability. Strong relationships are preferred over formalizing processes and enacting regulations because this soft law assists in compliance efforts. Strong relationships, however, cannot prevent corporate capture or pre-existing dynamics among actors who bolster detrimental "institutional logics" in processual accountability (Selbst, 2021). It is therefore not surprising that many researchers have foregrounded the limitations of the processual (Lee et al., 2019; Metcalf et al., 2021; Patil et al., 2014; Waldman, 2019), which were confirmed by our findings.

When governments indicate that they have documentation on "de-risking and mitigation measures" through AIAs but do not make these documents accessible to the public, governments fall short of true accountability. Transparency requires more than simply signaling the existence of such measures; it necessitates making the details publicly available to ensure meaningful oversight. Our findings resonate with Fox (2007) who argued that transparency does guarantee accountability. He introduced the concept of an "opaque transparency", in which "the dissemination of information… does not reveal how institutions actually behave in practice, whether in terms of how they make decisions, or the results of their actions" (Fox, 2007, p. 667). Even though the documents may be available, this tactic may be used by government to dilute a fulsome conceptualization of compliance.

In its design, Canada's AIA offers accountability. However, its results offer a list of check-box answers with little to no transparency or procedural fairness beyond the initial publications of the PDFs (and accompanying JSON files). This validates the consequences of pull-down menus and binary, yes/no boxes noted by Selbst (2021). Here transparency served to demonstrate that AI systems were designed and implemented "responsibly" in government.[16]Gertler (2023) asked whether the AIAs could affect change. He concluded that accountability in Canada's AIA was reduced to transparency and might instead be merely performative (Gertler, 2023). To further reduce the burden of compliance while increasing the performativity with AIAs, we suspect generative AI will increasingly be used to produce AIAs.

Canada's AIA focuses on conducting the assessment at the algorithmic design stage, and not at the deployment (use) stage, so the AIAs "cannot further educate regulators or the public as to the kinds of pressures, choices, and tradeoffs that the engineers and their managers must make in real practice [i.e., the outcomes]… it does not allow us to ask questions that we do not yet know to ask" (Selbst, 2021, p. 150). In other words, AI systems were developed under the assumption that, if they were developed with good intent, they would produce a beneficial outcome and the deployers could smoothly manage the implementation. The AIA process demands considerable public trust in government, a trust that is diminishing among the very liberal democracies that deploy AIAs. Absent that connection to real practice, that trust may continue slipping.

[16] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html

6 Conclusion

Canada's AIAs are not delivering as designed. In part, the gap between design and reality exists because, with AIAs, "algorithmic decision-making hides the fact that engineers and their corporate employers are choosing winners and losers while steadfastly remaining agnostic about the social, political, and economic consequences of their work" (Waldman, 2019, p. 616). Barring a rejection of AIAs or an overhaul of the underlying political and legal regime, we present the following recommendations.

AIA developers should reimagine accountability to include civil society, not as a target of communication and explanation, but as equal partners. This aligns with Bovens (2009, p. 199) who argued for a more direct and political accountability in which "agencies or individual public managers should feel obliged to account for their performance to the public at large or to civil interest groups, charities."

Canada is known for its risk-averse public sector. However, risk continues the processual mitigation of any harms, with a reframing of risk as reputational or financial loss (i.e., government funding). The AIA should move from risk to explicitly list harms (Metcalf et al., 2021). Government also should embrace a more nuanced understanding of internal and external drivers of AI systems, which we observed in our examination of the responses to the AIAs. AIAs could more closely tie harms to the actual experiences of affected communities, rather than casting impacts as abstract concerns or those amenable to internal metrics. Formalizing the characteristics of harms could serve to decrease definitional uncertainty.

We further recommend addressing the shortcomings of the design-reality gap, especially where entities feel no external pressure, other than intense public scrutiny, to comply with AIA mandates. Canada has experienced several high profile AI- and software-related scandals, for example with ArriveCAN (Curry, 2024), where lessons learned from failures appear to have failed to be implemented. Independent oversight is therefore critical, as self-regulation for AI has been found ineffective, especially when quasi-regulatory departments in charge of enacting legislation also have a role in promoting AI (Attard-Frost, 2023; Brandusescu & Sieber, 2025). Addressing the design-reality gap means broadening the humans-in-the-decision-making-loop. Creating external (i.e., second-party) tracking mechanisms and independent (i.e., third-party) oversight that blend the process and the outcome (e.g., Patil et al., 2014) is equally important. We hope these recommendations will retain the viability of the AIA, its design, and responses.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44206-025-00221-7.

Author contribution Both authors have contributed equally to the conceptualization, methodology and writing—original draft, writing—review and editing and approved the final submitted draft. The first author contributed in greater part to the data curation, formal analysis, project administration, and visualization.

Funding This research was supported by a Canada Graduate Scholar Doctoral program grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) (767-2021-0252) as well as a SSHRC Insight Grant (435-2023-0561). The funder had no role in study design, data collection and analysis, the decision to publish or preparation of the manuscript.

Data availability All resources used are included in references. Where they are available online, a link is provided. We created a repository comprised of two open datasets that we created and made publicly available at https://osf.io/rk8ux/.

Declarations

Conflict of interest There is no conflict of interest. While we have been asked to consult with TBS (the developers of the AIA), we have always done this at arms-length and TBS comes to us because we are critics. We are not paid as consultants nor employed by the government.

References

Ada Lovelace Institute. (2022). AIA User Guide. https://www.adalovelaceinstitute.org/resource/aia-user-guide/.

Ananny, M., Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, *20*(3), 973–989.

Ashar, A., Ginena, K., Cipollone, M., Barreto, R., & Cramer, H. (2024). Algorithmic impact assessments at scale: Practitioners' challenges and needs. *Journal of Online Trust and Safety*, *2*(4). https://doi.org/10.54501/jots.v2i4.206.

Attard-Frost, B. (2023). Generative AI systems: Impacts on artists & creators and related gaps in the artificial intelligence and data act. *Submission to the Standing committee on industry and technology*. https://www.ourcommons.ca/Content/Committee/441/INDU/Brief/BR12541028/br-external/AttardFrostBlair-e.pdf.

Attard-Frost, B., Brandusescu, A., & Lyons, K. (2024). The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. *Government Information Quarterly*, *41*(2), 101929. https://doi.org/10.1016/j.giq.2024.101929.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, *13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x.

Bovens, M. (2009). Public accountability. In E. Ferlie, L. L.E Jr, & C. Pollitt (Eds.), *The Oxford handbook of public management* (pp. 182–208). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199226443.003.0009.

Brandusescu, A. (2021). Artificial intelligence policy and funding in Canada: Public investments, private interests. In *Centre for interdisciplinary research on montreal*. Montreal, QC: McGill University. https://www.mcgill.ca/centre-montreal/files/centremontreal/aipolicyandfunding_report_updated_mar5.pdf

Brandusescu, A., & Sieber, R. E. (2025). Missed opportunities in AI regulation: Lessons from Canada's AI and data act. *Data & Policy*, *7*, e40. https://doi.org/10.1017/dap.2025.17.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080.

Centre, S. (2024). Tracking automated government 'TAG' register Canada. *Western university*. http://tagregistercanada.ca/.

Christian, G. (2024). CBSA border surveillance: The dangerous expansion of facial recognition technology. *Canadian immigration lawyers association*. https://cila.co/cbsa-border-surveillance-the-dangerous-expansion-of-facial-recognition-technology/.

Council of Europe. (2021). Human rights, democracy and rule of law impact assessment of AI systems. *ad hoc committee on artificial intelligence (cahai) policy development group (CAHAI-PDG)*. https://rm.coe.int/cahai-pdg-2021-02-subworkinggroup1-ai-impact-assessment-v1-2769-4229-7/1680a1bd2d.

Curry, B. (2024). Senior officials reject allegation that employee linked to ArriveCan was told to lie to investigators. *The globe and mail*. https://www.theglobeandmail.com/politics/article-arrivecan-employees-allegations-investigation.

Daly, P. (2023). Mapping artificial intelligence use in the government of Canada. *Revue Gouvernance/Governance Review*, *20*(1), 74–95. https://www.erudit.org/en/journals/gouvernance/2023-v20-n1-gouvernance08729/1106045ar/.

Darbyshire, T. (2022). In praise of the Canadian algorithmic impact assessment framework. *Tech UK*. https://www.techuk.org/resource/in-praise-of-the-canadian-algorithmic-impact-assessment-framework.html.
DeWalt, K. M, & DeWalt, B. R. (2010). *Participant observation: A guide for fieldworkers*. Maryland, US: Rowman & Littlefield.
Diakopoulos, N. (2020). Accountability, transparency, and algorithms. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 197–213). Oxford University Press.
Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., & Busch, C. (2020). Demographic bias in biometrics: A survey on an emerging challenge. *IEEE Transactions on Technology and Society*, *1*(2), 89–103.
European Commission. (2020). AI high-level expert group - assessment list for trustworthy artificial intelligence. https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-assessment-list-trustworthy-artificial-intelligence-altai?language=fr.
European Law Institute. (2022). Model rules on impact assessment of algorithmic decision-making systems used by public administration. https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Model_Rules_on_Impact_Assessment_of_ADMSs_Used_by_Public_Administration.pdf.
Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, *17*(4–5), 663–671. https://doi.org/10.1080/09614520701469955.
Gertler, N. (2023). *Hacking AI governance: Exploring the democratic potential of Canada's algorithmic impact assessment* [Masters thesis]. Concordia University. https://spectrum.library.concordia.ca/id/eprint/992742/. Montreal, Canada.
Gertler, N. (2024). Canada's algorithmic impact assessment. In F. McKelvey, S. Toupin, & J. Roberge (Eds.), *Northern lights and silicon dreams: AI governance in Canada (2011-2022)* (pp. 31–41). Montreal, Canada. https://www.amo-oma.ca/en/ai-policy-report/.
Gonzalez, V., & Salomon, G. (2025). Orders to leave the country - some for US citizens - sow confusion among immigrants. Associated Press. https://apnews.com/article/trump-immigration-cbp-one-asylum-4a3aae0453d17b6dfc2c9b590f19497e.
Government of Canada. (2022). *Algorithmic Impact Assessment*. Version 0.9.1. *Treasury Board of Canada Secretariat*. https://github.com/canada-ca/aia-eia-js/milestones.
Government of Canada. (2023). Directive on automated decision-making. *Treasury board of Canada secretariat*. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592#appC.
Government of Canada. (2024a). Algorithmic impact assessment tool. *Treasury board of Canada secretariat*. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.
Government of Canada. (2024b). Algorithmic impact assessment version 0.10.0. *Treasury board of Canada secretariat*. https://canada-ca.github.io/aia-eia-js/.
Government of Canada. (2024c). Guide on the scope of the directive on automated decision-making. *Treasury board of Canada secretariat*. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-scope-directive-automated-decision-making.html.
Government of Canada. (2025). *Open Government Portal: "Algorithmic Impact Assessments"*. Treasury Board of Canada Secretariat. https://search.open.canada.ca/opendata/?sort=metadata_created+desc%26search_text=%22algorithmic+impact+assessment%22%26;page=1.
Hasan, A., Brown, S., Davidovic, J., Lange, B., & Regan, M. (2022). Algorithmic bias and risk assessments: Lessons from practice. *Digital Society*, *1*(2), 14. https://doi.org/10.1007/s44206-022-00017-z.
Heisler, N. (2022). *Standards for the control of algorithmic bias in the Canadian administrative context* (Master's thesis, University of Waterloo). https://uwspace.uwaterloo.ca/items/4939baca-d81c-4531-934d-83e22c83cd49. Waterloo, Canada.
Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *The Qualitative Health Research*, *15*(9), 1277–1288.
Institute for the Future of Work. (2022). International learnings on algorithmic impact assessments (panel). https://www.youtube.com/watch?v=kFti_oKZFsw.
Institute for the Future of Work. (2023). Good Work AIA. https://www.ifow.org/publications/good-work-algorithmic-impact-assessment-an-approach-for-worker-involvement.
Karadeglija, A. (2024). Federal government use of AI in hundreds of initiatives revealed by new research database. *CBC news*. https://www.cbc.ca/news/politics/federal-government-used-ai-1.7170307.

Karlin, M. (2018). Deploying AI responsibly in government. *Policy options*. https://policyoptions.irpp.org/magazines/february-2018/deploying-ai-responsibly-in-government/.

Kaye, K. (2024). How Canada's algorithmic impact assessment process and algorithm has evolved. *World privacy forum*. Retrieved from https://www.worldprivacyforum.org/wp-content/uploads/2024/08/WPF_AI_Governance_Canada_AIA_August2024_fs.pdf.

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

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, K. (2019, November). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM Human-Computer Interaction*, *3*(182), 26 pp. https://doi.org/10.1145/3359284.

Luna-Reyes, L. F., Pardo, T. A., Gil-Garcia, J. R., Navarrete, C., Zhang, J., & Mellouli, S. (2010). Digital government in North America: A comparative analysis of policy and program priorities in Canada, Mexico, and the United States. In C. Reddick (Ed.), *Comparative E-Government* (pp. 139–160). Integrated Series in Information Systems, 25. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-6536-3_7.

Metcalf, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021, March). Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM Conference On Fairness, Accountability, and Transparency* (pp. 735–746). https://dl.acm.org/doi/abs/10.1145/3442188.3445935.

Mikhaylov, S. J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170357. https://doi.org/10.1098/rsta.2017.0357.

Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines*, *31*, 323–327.

Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021). Assembling accountability: Algorithmic impact assessment for the public interest. *Data & society research institute*. https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/.

Mulligan, D. K., & Bamberger, K. A. (2019). Procurement as policy: Administrative process for machine learning. *Berkeley Technology Law Journal*, *34*, 773.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

Patil, S. V., Vieider, F., & Tetlock, P. E. (2014). Process versus outcome accountability. In M. A. P. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook for public accountability* (pp. 69–89). https://doi.org/10.1093/oxfordhb/9780199641253.013.0002.

Public Policy Forum. (1998). Innovation in the federal government: The risk not taken. *background document for a roundtable discussion to be held on behalf of the office of the Auditor General*. https://www.oag-bvg.gc.ca/internet/english/meth_gde_e_10193.html.

Quenneville, G. (2022). Meet the librarian in rural Vermont helping stranded travellers with the Arrive-CAN app. *CBC news*. https://www.cbc.ca/news/politics/vermont-librarian-arrivecan-app-1.6562293/.

Reeveley, D. (2021). Federal rules on AI too narrow and risk 'damaging public trust': Internal review. *The logic*. https://thelogic.co/news/federal-rules-on-ai-too-narrow-and-risk-damaging-public-trust-internal-review.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. *AI Now Institute*. https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2.

Sambuli, N. (2021). Five challenges with multistakeholder initiatives on AI. In *Artificial intelligence and equality initiative*. *Carnegie Council for Ethics and International Affairs*. https://www.carnegiecouncil.org/media/article/five-challenges-with-multistakeholder-initiatives-on-ai.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry. Preconference at the 64th Annual Meeting of the International Communication Association*, Seattle, WA, USA. https://websites.umich.edu/%7Ecsandvig/research/Auditing%20Algorithms%20-%20Sandvig%20-%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.

Schneider, K. (2024). RCMP's proposed AI surveillance system for holding cells called 'dehumanizing' and 'intrusive' by experts. *Investigative journalism foundation*. https://theijf.org/rcmp-ai-surveillance-system.

Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. *Harvard Journal of Law & Technology (Harvard JOLT)*, *35*(1), 117–192. https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf.

Sieber, R. E. (2022). Raw political power in civic engagement with AI. In A. Brandusescu & J. Reia (Eds.), *Artificial intelligence in the city: Building civic engagement and public trust* (pp. 17–18). https://doi.org/10.18130/9kar-xn17.

Sieber, R. E., Brandusescu, A., Adu-Daako, A., & Sangiambut, S. (2024). Who are the publics engaging in AI? *Public Understanding of Science*, *33*(5), 634–653. https://doi.org/10.1177/09636625231219853.

Sieber, R., Brandusescu, A., Sangiambut, S., & Adu-Daako, A. (2025). What is civic participation in artificial intelligence? *Environment and Planning B: Urban Analytics and City Science, 52*(6), 1388–1406. https://doi.org/10.1177/23998083241296200.

Sloane, M., & Moss, E. (2023). Assessing the assessment: Comparing algorithmic impact assessments and AI audits. *SSRN*, *14*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4486259.

Smith, L. G. (1984). Public participation in policy making: The state-of-the-art in Canada. *Geoforum*, *2*, 253–259.

Stahl, B. C., Antoniou, J., Bhalla, N. B., Jansen, L., Lindqvist, P., Kirichenko, B., Marchal, A., Rodrigues, S., Santiago, R., Warso, N., Z, & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, *56*, 12799–12831. https://doi.org/10.1007/s10462-023-10420-8.

Waldman, A., & Martin, K. (2022). Governing algorithmic decisions: The role of decision importance and governance on perceived legitimacy of algorithmic decisions. *Big Data & Society*, *9*(1), 16 pp. https://doi.org/10.1177/20539517221100449.

Waldman, A. E. (2019). Power, process, and automated decision-making. *Fordham Law Review*, *88*, 613. https://ir.lawnet.fordham.edu/flr/vol88/iss2/9/.

Watkins, E. A., Moss, E., Metcalf, J., Singh, R., & Elish, M. C. (2021). Governing algorithmic systems with impact assessments: Six observations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 1010–1022). New York City, NY, USA. https://doi.org/10.1145/3461702.3462580.

Wernick, A. (2024). Impact assessment as a legal design pattern-A "timeless way" of managing future risks? *Digital Society*, *3*(29), 36 pp.

Wieringa, M. (2020, January). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18).