

24 MAY 2023 • 21 MINUTE READ

SHARE

How artificial intelligence will change administrative law: The Government of Canada's Directive on automated decision-making

Written by: [Shea Coulson](#)

Contributors: Trevor Hunt (Articling Student)

Most media attention placed on artificial intelligence today focuses on the efficiency gains and impacts on the job-market, the implications for privacy, data protection, or the commercial risks created by AI-assisted decision making. However, the impact of artificial intelligence on administrative decisions by governments and governmental agencies is equally profound and deserves more attention.

This article will review newly updated federal policy, canvass the administrative law issues raised by AI systems, and review some U.S. case law where AI systems have been considered in court challenges of administrative decisions as a prelude to the kind of cases we will start to see in Canada.

Governments in the United States, the EU, the UK and Canada are now using AI and algorithmic systems to assist in decision making for a wide range of purposes including government benefits, immigration, environmental risk assessment, economic regulation, and assisting in regulatory investigations. These systems are helping governments improve efficiency, reduce costs, and improve performance metrics. While certain automated systems have been in place for years, the changing landscape in AI, including the explosion of generative AI systems, is creating new risk for governments. Accordingly, the use of these technologies is creating new issues in administrative law that are [generating significant debate and discussion internationally](#) and that require a major rethinking of administrative decision making and effective means to engage with and challenge such decisions.

In April, 2023, the federal government updated its 2020 "Directive on Automated Decision-Making" (the "Directive") following a stakeholder review of the Directive to re-evaluate and adapt the state and use of the Directive based on the current Canadian and global AI landscape, and the evolving risks created by these changes. The Directive is the first national policy focused on algorithmic and automated decision-making in public administration. It tells us a lot about how government is understanding the impact of AI systems on decision making and highlights some of the administrative law problems created by their use. The government developed the Directive based on AI guiding principles adopted by leading digital nations at the [D9 conference in November 2018](#). The principles include: (1) understanding and measuring the impact of using AI; (2) ensuring transparency about how and when AI is used; (3) providing meaningful explanations about AI decision-making; (4) remaining open to sharing source codes and relevant data while protecting personal data; and (5) providing sufficient training for the use and development of AI solutions.

The Directive is part of the federal government's overarching [Policy on Service and Digital](#), which forms the national strategy on digital government, including its AI strategy. There is no legislation that currently regulates the use of AI and algorithmic systems by government. These soft-law instruments therefore only have legal force insofar as they inform common law administrative law obligations although it should be noted that various agencies may have directives about AI use in orders in council or permissive legislative authorizations.

Federal policy on automated decision making

The Government of Canada has recognized the risk of automated decisions (including those assisted by simpler linear models and by complex AI models), by promulgating a policy directive on Automated Decision-Making under the [Financial Administration Act](#) and the [Public Service Employment Act](#). The Directive's objectives are:

- Decisions made by federal institutions are data-driven, responsible and comply with procedural fairness and due process requirements.
- Impacts of algorithms on administrative decisions are assessed and negative outcomes are reduced, when encountered.
- Data and information on the use of automated decision systems in federal institutions are made available to the public, where appropriate.

The Directive applies as follows:

- to any system, tool, or statistical model used to make an administrative decision or a related assessment about a client.
- to automated decision systems in production and excludes systems operating in test environments.
- to all institutions subject to the Policy on Service and Digital, which is effectively all departments listed in Schedule I of the [Financial Administration Act](#), R.S.C. 1985, c. F-11 (E.g. Department of Natural Resources, Department of Transport; Department of Health; Department of Employment and Social Development).

Given this breadth, the Directive is clearly intended to capture a wide-array of administrative bodies and signals the federal government's view that automated decision making has profound impacts on the administrative law requirements to which all administrative bodies must adhere. That said, there are still many bodies that are not captured by the Directive, and these gaps should be reduced and ultimately eliminated. There is no administrative decision that

ought to be exempt from careful requirements intended to safeguard procedural fairness and the quality of substantive decisions where AI and algorithmic systems are deployed. Provincial governments also need to develop similar directives for decision making bodies under provincial jurisdiction. There are provincial governments with AI strategies, but nothing as comprehensive as the federal Directive. See, for example, [Ontario's Trustworthy Artificial Intelligence Framework Consultation](#) and [Beta principles for the ethical use of AI and data enhanced technologies in Ontario](#); in B.C. the provincial government has a set of "[Digital Principles](#)" to guide the work of public service employees, which touch on artificial intelligence, but provide little meaningful guidance.

The Directive imposes positive obligations on Assistant Deputy Ministers or others legally responsible for the program using the automated decision system, including:

- Algorithmic impact assessment
- Transparency
- Quality assurance
- Recourse
- Reporting

The Directive also requires the Chief Information Officer of Canada to provide government-wide guidance on the use of automated decision systems, develop and maintain the algorithmic impact assessment, and develop common strategies, approaches, and processes to support the responsible use of automated decision systems.

Algorithmic impact assessment

The program head will need to apply an [algorithmic impact assessment](#) ("AIA") prior to the production of any automated decision system. This assessment focuses on both risk assessment and mitigation and is intended to assist agencies in risk assessment using a series of standard questions, a scoring system to assess risk creation and risk mitigation to determine the level of impact (Level I (little to no impact) to Level IV (very high impact) of each automated system, focusing particularly on:

- The rights of individuals or communities;
- The health or well-being of individuals or communities;
- The economic interests of individuals, entities, or communities; and
- The ongoing sustainability of an ecosystem.

In addition to assessing individual impacts, the AIA includes sections on algorithmic risk, the data source, data quality, procedural fairness, and privacy.

Government legal services must also be consulted as part of the risk assessment process. The higher the risk identified by the relevant agency the stricter the required safeguards to ensure procedural fairness. For example, risk levels II and III require some form of peer review by qualified experts and level IV risks require robust peer review by qualified experts.

While the AIA is a useful tool for government bodies, it cannot in itself guarantee a procedurally fair result. As technology rapidly evolves, the AIA will need to be updated. Unlike simpler, linear models, complex AI models such as generative AI pose significant challenges under the current AIA. For example, it can be notoriously difficult to properly identify the impact of a generative AI system (or even be aware of its use) within the administrative process. It is also challenging to assess and explain how the algorithms that drive these systems arrive at the outputs that they do (which is currently a required step in the AIA process). These limitations mean the risks these systems can create and the biases they can introduce may not be easily detectable or traceable back to the AI system. This could lead to a risk assessment lower than the actual risks created, which in turn could lead to insufficient (and reviewable) procedural fairness safeguards for a particular decision.

Arguably this risk assessment could properly form part of the record for a judicial review of any decision made or assisted by an automated decision system that fits within the Directive. The updates to the Directive require the results of each assessment to be released on the Open Government Portal so the AIAs will be accessible. While imperfect, by positively requiring this assessment prior to the deployment of an automated decision system, the federal government is again signalling that these systems can have profound impacts on the fairness and reasonableness of administrative decisions.

To date, AIAs have been conducted on temporary resident visa and work permit application processes, tools to enhance immigration fraud assessment, tools to assess privately sponsored refugee applications, systems that determine eligibility of spouse and common law partners for immigration visas, tools to assess mental health benefits for veterans, and the use of natural language models to automate the review of free text comments received on records of employment in the unemployment benefits process. This limited set of AIAs shows that the risk assessment requirement is insufficiently capturing the use of AI and other automated systems by government decision makers, which extends far beyond the above. This does not mean that the use of systems for which there are no AIAs are exempt from the procedural fairness and substantive law obligations reflected in the principles underlying the AIA process set by the Directive.

Transparency

Procedural fairness dictates that all those impacted by administrative decisions must be given proper notice of those decisions and an opportunity to be heard. The transparency portion of the objective is intended to satisfy these obligations. It specifically requires:

- Providing notice before decisions. Higher risk levels require more robust notice (e.g. risk impact Levels III and IV under the AIA require plain language notice to impacted parties);
- Providing explanations after decisions (how and why the decision was made). Note here that AIA impact levels I and II do not require human involvement in the decisions but impact levels III and IV do);
- Access to components and release of source code: i.e. allowing access to the instructions that led to the decision or the assisted decision; and
- Documenting decisions.

These requirements impose an obligation on a decision-maker to have the ability to explain how an automated decision system works in relation to the decision being rendered and the impact it has on that decision. The intention is for those impacted by such decisions to have an opportunity to properly review and understand them. Of course, this ability is relative to the impacted party having sufficient expertise in how automated decision systems work.

However, in reality it can be challenging to know when an algorithm has been used as part of a decision making process, particularly for pre-existing systems to which the Directive does not apply. Partial disclosure is also common, with some systems being disclosed and others not. The federal government has recognized these issues in updating its directive to be more encompassing in scope, but continued improvements in transparency and notice will be essential to ensuring robust procedural fairness and protections for those impacted by decisions assisted or made by AI systems.

Quality assurance

The Directive imposes quality assurance measures to mitigate the risks of automated decision systems:

1. Testing and monitoring outcomes: the data and information used by the automated decision system and the underlying model must be tested for bias and unintentional outcomes prior to deployment.
2. Data quality: Data collected for, and use by, the automated decision system must be validated as relevant, accurate, up-to-date, and in accordance with the Policy on Service and Digital and the Privacy Act.
3. Data governance: establishing measures to ensure that the data used and generated by automated decision systems are traceable, protected and accessed appropriately, and lawfully.
4. Peer review: consulting qualified experts to review automated decision systems.
5. Gender-based Analysis Plus: these assessments must be completed during the development or modification of an automated decision system.
6. Employee training: providing adequate employee training in the design, function, and implementation of the automated decision system to be able to review, explain and oversee its operations.
7. IT and business continuity management: strategies and plans to support IT and business continuity management.
8. Security: security assessments will be part of the AIA.
9. Legal: consultation with the institution's legal services from the concept stage of an automation project to ensure that the use of the automated decision system is compliant with applicable legal requirements.
10. Ensuring human intervention: As noted above, certain AIA risk levels will require human intervention in decisions.

These principles are useful and, if perfected, would greatly reduce issues. However, experience reveals that there are many areas where quality breaks down in practice. For example, data quality is often a significant issue for many organizations and is likely to pose significant issues for government systems as well. Poor data quality could lead to completely incorrect statistical analyses, correlation issues, misleading data, and faulty outcomes. Without robust safeguards, data can also easily be manipulated (intentionally or not) to provide distorted perspectives on issues - and this can be the case both for decision makers and parties who have vested interests in achieving particular outcomes in the regulatory process.

Employee training will be crucially important to protect against misleading uses of data, incorrect inferences and deductions, and effective vetting and oversight. The higher AIA impact risk of a system, the greater the training required of those who use it.

Recourse and reporting

The Directive requires institutions to provide clients with recourse to challenge administrative decisions. The Directive also requires publication of information on the effectiveness and efficiency of the automated decision system in meeting program objectives on a government website.

Concluding thoughts on the Directive

While the Directive is a useful starting point, the impact of AI-assisted decision making on both procedural fairness and the substantive aspects of a decision is untested in Canada. The Directive clarifies many of the key areas for judicial scrutiny and will be a useful guide for those wishing to understand how to interrogate decisions made or assisted by automated decision systems such as AI systems. However, there is little doubt these systems will introduce entirely new categories of risk in administrative decisions and thereby create fruitful grounds for review.

Lessons from US case law

Canadian jurisprudence

There is no case law in Canada addressing the use of AI systems in government decisions. However, the Supreme Court of Canada did consider principles that will likely apply to the use of AI in decision making in *May v. Ferndale Institution*, [2005] 3 SCR 809 ("*Ferndale*"). In *Ferndale* the court assessed the use of a scoring matrix to classify prisoners into medium-security versus minimum-security prisons. The court held that it was a breach of procedural fairness not to disclose details of the underlying components of the scoring matrix based on classic administrative law principles requiring notice of the relevant factors to a decision - indicating that similar reasoning could apply to decisions made or assisted by AI or algorithmic systems. In other words, the underlying components of these systems and how they impact the ultimate decision will likely have to be disclosed in some fashion to satisfy procedural fairness obligations.

On top of the obvious procedural fairness concerns, it is hard to understand how the *Vavilov* principles on substantive review will dovetail with AI systems. *Vavilov* imposes principles of justification and transparency that require that an administrative decision maker's reasons meaningfully account for the central issues and concerns raised by the parties. The requirement for "responsive reasons" that are transparent and can be reviewed by courts imposes serious challenges on decision makers using AI or algorithmic systems that produce outputs using models that cannot easily be explained.

The Directive attempts to address this issue by imposing an explanation requirement relative to the AIA impact assessment level. For example, level I systems can render decisions that are explained by a FAQ that provides details for common decision results. Impact levels II, III, and IV require a greater explanation for any decision that results in the denial of a benefit, service, or other regulatory action. This poses challenges in converting technical bases for decisions to explanations that are both understandable and that can also be interrogated. The extent to which plain language or simplification will be required is unknown at this point but the Directive distinguishes between interpretability and explainability.

Interpretability means that a human can easily discern the prediction made by the model directly from the inputs. Explainability means providing a human with a set of techniques to explain a prediction made by an automated decision system but not necessarily providing the model itself. Clearly there is a world of difference between basic linear models and complex predictive neural networks and systems such as large language models. The quality and sufficiency of such explanations will be tested under the *Vavilov* principles and it is not yet clear what kind of guidance Canadian courts will provide for different types of AI and algorithmic systems to satisfy justification requirements. It is also questionable whether courts are properly equipped to assess the reasonableness of an AI or other model or to properly adjudicate expert evidence on these models. Significant challenges lie ahead.

The United States courts have seen quite a few more challenges involving artificial intelligence, and provide useful guidance for Canada.

Unemployment insurance

In *Cahoo v. SAS Analytics Inc.*, 912 F.3d 887 (6th Cir. 2019), the Michigan Unemployment Insurance Agency (MUIA) used an artificial intelligence system called MiDAS to administer its unemployment benefits. "When MiDAS detected unreported income or "flagged" other information about a claimant, it initiated an automated process to determine whether the individual had engaged in fraudulent behavior." When an individual was flagged for possible fraud, a multiple-choice questionnaire was sent to them, and any affirmative answer to a question resulted in the AI deeming the individual's actions as being fraudulent. There was no communication to individuals about why they were under suspicion or how to rebut the allegation, and additionally these determinations of fraud were made exclusively by MiDAS with no human involvement from October 2013 to August 2015.

Individuals found to have engaged in fraudulent behavior by MiDAS had their unemployment benefits terminated and were automatically given fines of the maximum penalty allowed under state law. The MUIA automatically sent letters to

These automatically generated letters of administrative penalty advised that state law required the imposition of secondary penalties on the individuals that notified them of the determination of fraud and stated secondary penalties to be imposed if the initial fines weren't paid. These letters did not state the factual basis for a finding of fraud. The Michigan Auditor General reviewed 22,000 of MIDAS' fraud determinations and found a 93 percent false positive rate.

The Michigan 6th Circuit Court of Appeals quashed the decision and found violations of procedural fairness because the AI system did not provide sufficient notice and opportunity to be heard appropriate to the nature of the case. Though this case was argued under the due process clause of the U.S. constitution, the principles are parallel to Canadian doctrines of procedural fairness required in administrative decision making. Similarly see, *Sterling v. Feek*, 2022 U.S. Dist. LEXIS 200758, a pre-trial motion in a case involving a claim that Washington's reassessment of employment benefits using an automated system violated due process rights.

Criminal sentencing

In *State v. Loomis*, 2016 WI 68, Eric Loomis was sentenced to 11 years in prison on charges relating to a drive-by shooting. During his initial sentencing hearing, the court considered a COMPAS risk assessment contained within his Presentence Investigation Report (PSI) which provides separate scores for pretrial recidivism risk, general recidivism risk, and violent recidivism risk. Loomis had received high risk scores in each category.

While not a product of generative AI, COMPAS applies algorithmic predictions using both qualitative and quantitative factors to generate a risk score for each category, based on population data that exist within its data sets. A non-exhaustive list of factors considered by COMPAS in making its predictions is as follows:

- Pretrial risk: current charges, pending charges, prior arrest history, previous pretrial failure, residential stability, employment status, community ties, and substance abuse.
- General recidivism: prior criminal history, criminal associates, drug involvement, and early indicators of juvenile delinquency problems.
- Violent recidivism: history of violence, history of non-compliance, vocational/educational problems, the person's age-at-intake and the person's age at first arrest.

Loomis appealed the sentencing decision and argued that the circuit court's decision violated his due process rights under the 14th Amendment by considering the COMPAS risk assessment in his sentence. The circuit court upheld the original sentence and the Wisconsin Supreme Court (WSC) dismissed the appeal, finding that the use of the automated system did not violate Loomis' due process rights, particularly because the COMPAS assessment was not the determinative factor in the impugned. In that decision the court provided interesting guidance on assessing automated systems such as COMPAS: (1) the risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations; (2) the COMPAS score cannot be considered an aggravating or mitigating factor in sentence length, but can only inform the manner in which the sentence is to be carried out; (3) the system scores should be used only for the purpose for which they were intended, which is to reduce the risk of recidivism. This is only one of many factors a judge considered when making a sentencing decision.

Teacher evaluations and terminations

In *Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist.*, 251 F. Supp. 3d 1168 (S.D. Tex. 2017), the Houston Independent School District (HISD) used SAS' EVAAS software to generate scores of teacher effectiveness derived from improvements to their students' grades as compared to other teachers statewide. The HISD then used the scores as a basis for the termination of teachers with low scores. The Houston Federation of Teachers (HFT) filed a lawsuit against HISD under the 14th amendment for violation of due process rights.

HISD sought summary judgment to dismiss the claims. The court dismissed the procedural fairness aspects of the application because EVAAS scores could not be independently verified and there was a risk of coding errors that terminated teachers would not be able to identify and appeal, despite these errors contributing to their termination. Where errors could be identified, the nature of the algorithm meant that the recalculation of a teacher's individual score could potentially change scores across the entire district or state because the scores were relative to one another.

Final thoughts on US cases

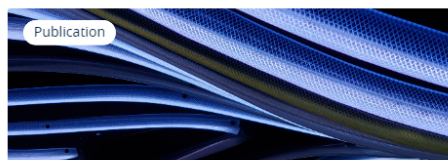
The US cases demonstrate that the issues identified in the Directive and its AIA as areas of risk do manifest in administrative decisions and can radically skew outcomes. The explainability of AI systems is relevant to procedural fairness as is the ability of an impacted person to interrogate and respond to the basis for the AI decision, which includes an ability to challenge coding errors and algorithmic bias.

What does this mean for regulated industries?

While each regulatory body has its own specialized remit and expertise, the impact of automated decision making will cross numerous sectors and will create common risks within administrative decision making. Regulatory counsel will need to understand how any automated decision making system works, exactly how any algorithms or statistical models impact administrative reasoning and decision-making processes, and use that knowledge to uncover issues with respect to bias, data accuracy, decision quality, and procedural fairness issues such as transparency, the opportunity to be heard, and notice requirements.

This is no small task. Unfortunately, there is currently no regulatory oversight body that can assist in ensuring AI fairness in government decision making, despite there being [calls for this kind of oversight](#). Instead, industries impacted by government decisions using AI and algorithmic systems will need to develop their own expertise in responding to the impact of these systems on regulatory decision making. It is early days, but there is no doubt these systems will profoundly remake the administrative and regulatory state.

Related insights



Publication

Using policy to protect your organization from generative AI risks

8 MAY 2023 · 19 MINUTE READ

Related capabilities

- Technology
- Regulatory and Government Affairs
- Data Protection, Privacy and Cybersecurity